UNIVERSITÉ D'ORLÉANS

ÉCOLE DOCTORALE

Énergie - Matériaux - Sciences de la Terre et de l'Univers

Institut des Science de la Terre d'Orléans

THÈSE présentée par :

Imadeddine LAOUICI

soutenue le : **17 Décembre 2024**

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Geological knowledge formalization and automation of the structural interpretation process for building 3D architectures of the sub-surface

Discipline/ Spécialité : Sciences de l'Univers / Modélisation

THÈSE dirigée par :

Mr Laurent ARBARET	MCF HDR, Université d'Orléans (directeur)
Mr Yannick BRANQUET	MCF HDR, Université de Rennes (co-directeur)
Mme Christelle LOISELET	Ing de recherche, BRGM (co-encadrante)
Mr Gautier LAURENT	MCF, Université d'Orléans (co-encadrant)

RAPPORTEURS :

Mr Guillaume CAUMON	Professeur, Université de Lorraine
Mme Mara ABEL	Professeur, Université fédérale du Rio Grande do Sul

---

EXAMINATEURS :

Mme Christel VRAIN	Professeur, Université d'Orléans, Président du jury
Mr Charles GUMIAUX	MCF, Université d'Orléans

INVITES :

Mr Boyan BRODARIC	Ing de recherche, service géologique national Canadien

Formalisation des connaissances géologiques et automatisation du processus d'interprétation structurale pour la construction d'architectures géologiques 3D du sous-sol

If they were:

Socrates, a structural geologist: *"**I know that I know nothing** about this outcrop, but I do know that an older fault cannot cut a younger unit"*

Avicenna, a determinist geologist: *"**The knowledge of anything**, even a minor fold**, is not complete unless we understand its causes**"*

Kant, a geophysicist: *"**All our knowledge begins with the senses, proceeds to understanding, and ends with reason**, except in subsurface exploration, where it ends with uncertainty"*

Confucius, a geomodeller: *"**Real knowledge is knowing the extent of one's ignorance** about the subsurface architectures beyond measured observations"*

Nietzsche, another geomodeller: *"In a geomodel**, there are no facts, only interpretations**"*

Remerciement

Ainsi, puisque, ‘quel que soit le savoir d’un homme, il en existe toujours un plus grand', cette thèse est la tentative obstinée d’un ignorant d’en savoir un peu plus… et d’exprimer toute sa gratitude envers ceux qui ont eu la patience de l’accompagner dans cette quête.

Arrivé à l’ISTO/BRGM avec des notions basiques de la géol 3D, un grand sac de motivation et zéro idée de ce qu’est une ontologie, j’ai dû me frotter à un monde foisonnant de concepts : géomodélisation, ontologies, sciences cognitives, modélisation conceptuelle, représentation des connaissances et intelligence artificielle. Peu à peu, entre questionnements et tâtonnements, j’ai fini par apprivoiser cet univers et me frayer un chemin pour devenir le *‘modélisateur* × *(connaissance, geol)’* que je suis.

Je commence par remercier Gautier, mon encadrant, pour m’avoir offert l’opportunité d’évoluer sous son égide. Tout au long de cette aventure, il m’a accordé la liberté de développer mes idées, tout en m’accompagnant dans les itérations de ce périple. J’ai une immense admiration pour son art de transmettre du complexe avec une clarté désarmante et pour son courage à sortir du ‘code’ et plonger tête la première dans des débats ‘ontologiques’ interminables –les derniers de la cantine. En discutant les origines des capes, les failles qui n’existent pas … et j’en passe, Gautier a su enrichir mon parcours non seulement sur le plan scientifique, mais aussi ‘Sémantique’, en m’emmenant explorer les confins du Robert. Je remercie également Christelle, mon encadrante au BRGM, pour son soutien indéfectible. Sa patience et sa vision du métier ont été le coup de pédale qui m’a permis de reprendre l’équilibre après les ‘chuteS’. Yvan-Eddine lui adresse sa reconnaissance pour l’avoir chaleureusement intégrée au sein du BRGM, et pour avoir partagé un espace de travail où l'échange et la collaboration ont été essentiels.

Un merci appuyé va à Yannick, mon directeur de thèse, pour ses conseils géo-logiques sur les terrains bretons, ariégeois, mais aussi orléanais, qui m’ont souvent permis de garder la tête hors de l’eau dans cette mer de complexité qu’est une thèse. Je remercie aussi Laurent Arbaret, officiellement directeur de thèse après le départ de Yannick, pour avoir accepté de reprendre le flambeau et pour son exposé sur les écrevisses rouges du lac Filheit. Enfin, je ne peux pas oublier Boyan, qui, bien que ne faisant pas partie de l’équipe officielle d’encadrement, n’a cessé d’apporter moult sagesse et aide précieuse tout au long de cette épopée. Sans son aide et sa patience infinie, naviguer en ces eaux troubles de l’atlontologie ‘got it !’, eût été un défi autrement plus rude. Et oui à la fin, j’ai pu traverser l’atlantique simplement avec une queue de castor (private joke) …

Un immense merci aux membres du jury – Christelle, Charles, Mara et Guillaume – pour avoir pris le temps de lire et évaluer ce travail. Franchement, si je pouvais refaire la soutenance, je prolongerais volontiers nos discussions bien au-delà de ces deux petites heures passées bien trop vite. Un merci spécial à mes rapporteurs, Mara et Guillaume, pour vos retours détaillés et vos conseils qui m’ont permis de peaufiner le manuscrit. Mara, tu as été là dès le début, depuis Tarbes, Enschede jusqu’à la dernière ligne droite à Orléans. Et oui, je m’excuse encore pour cette fameuse fois où je t’ai, bien involontairement, fait rater une session après un voyage transatlantique –mais que veux-tu, ton enthousiasme pour expliquer comment construire une bonne ontologie nous a embarqués dans des discussions sans fin. Guillaume, je dois te l’avouer : proposer paradigme entièrement différent pour la construction des systèmes de géomod, ça m’a longtemps donné des sueurs froides, malgré les résultats encourageants. Je me demandais constamment comment cela allait être perçu par des piliers de la discipline. Mais après avoir eu ton retour positif, j’ai eu un vrai déclic et ça m’a donné une confiance d’acier dans ce travail.

Je tiens également à exprimer ma gratitude aux financeurs de ce projet de thèse pour m'offrir l'opportunité d'explorer de nouvelles perspectives et de relever des défis passionnants ; le BRGM (en particulier Mathew), l'UO et l'ISTO (en particulier Lionnel) et l'ANR.

Au-delà de mes collaborateurs directs, ce boulot n'aurait jamais vu le jour sans tous les échanges et discussions interminables avec mes collègues du BRGM et de l'université. Merci aux deux Nicolas de l'équipe TIA pour ces discussions bien intenses sur la géostate, le ML, les scies, le russe… et j'en passe! Un clin d'œil à mon premier coloc au BRGM, Florent « le sacré chat », et à Musaab « le fameux match ». Fatma, merci aussi, même si tu es arrivée sur la fin, ton expertise a clairement été précieuse. Et bien sûr, un big up au reste de l'équipe : Hugo, Amaya, Lynn, Nathan, Djihad, Theo et Célia.

Un petit flashback à l'ancienne équipe DATA : merci à Sylvain et Corentin pour m'avoir mis dans le bain au tout début. Arnaud et Yann, Myriam merci pour vos encouragements ! Et Guillaume, bon courage pour le début de ton projet de recherche : je sens qu'on n'a pas fini de débattre ! Je ne peux pas oublier l'équipe semi-croustillante (on peut être fiers de notre victoire au Futuroscope)

À la clique de l'université, merci pour tous ces « petits » moments passés ensemble : François, Mahamadou, Tom, Thibault, Walid, Léo, bon courage pour la suite ! Aux anciens : Julien, Fabien, Valentin, vous m'avez bien accueilli au début, bonne continuation dans vos carrières ! Mention à Rabie, Ismael, Laurez (1/5, oui oui, je compte), Abdoul, et Rémi, qui va hériter de mon local (faut qu'on reparle du vélo). Merci aussi à Olivier, Marie-France, Fabienne, Nathalie, Julien, Yohan et Ferly pour votre aide quotidienne, que ce soit sur les galères admins, informatiques ou même les bricoles de dimanches.

Cette thèse, c'était aussi l'occasion de croiser une belle brochette de chercheurs hyper calés, qui m'ont bien aidé à y voir plus clair sur mon sujet : Anaïs, merci pour ton coup de pouce au tout début pour comprendre ce délire de triplets, merci aux gens de la 3D (Jean-François, Michel, Michael et Erik) pour les discussions et vos orientations, ceux de la websem (Jean-Jacques et Beatrice) et aux pros des onto (Emilio et Stefano).

Sans oublier d'où je viens et par où je suis passé, je tiens à remercier tous mes profs des 'nombreuses' universités que j'ai eu la chance de fréquenter. Vous avez tous contribué à me pousser à continuer tout droit, même quand la route semblait sinueuse. Particulièrement, merci Jean-Yves de Lille, Philippe de Bordeaux, Svetlana-Zorina et Vladmir-Vladimirovich de Kazan, ainsi que Fateh et Azzedine de Jijel. Un merci tout particulier à Bernard et Yvan anciennement du BRGM, qui ont cru en moi (y'a 6 ans). Grâce à vous, j'ai pu faire le saut de la géol 2D au 3D, un véritable tournant pour moi !

Et maintenant, place aux proches. Ce travail, je l'offre à mes parents, qui m'ont toujours soutenu depuis le tout début et qui ont planté en moi cette 'boulimie' pour la science et la recherche. À ma petite sœur et son chaton : « Grandi, ton-ton! ». À mes potes éparpillés aux quatre coins du monde : Aissa, Zizou, Silouh, Nb, Bilal, Amine, Ali, Raouf, Lamine, Alilou, Dob, Dja, Ya'k, Riad, Bouss, Ismail, les couz Trididay, Lilou, Bacha, Bilou, Saj, Ouss : on en a fait du chemin ! Bratan Moh, t'avais raison : chacun d'entre nous a fini par avoir ce qu'il a venu chercher, au bon moment ! Badia et ta petite famille (surtout rwijal), merci à vous tous, votre soutien moral a été un vrai pilier.

Et enfin… un merci tout spécial à celle sans qui tout ça serait tombé à l'eau. Tu m'as épaulé du début à la fin, Lmra, merci pour TOUT.

Dans l'espoir de n'avoir oublié personne (j'en oublie sûrement…), merci à vous tous !

Résumé

La modélisation structurale tridimensionnelle est un outil essentiel en géologie pour représenter l'architecture du sous-sol, faciliter la communication, améliorer la compréhension des experts et finalement prendre des décisions sur l'exploitation et la gestion des géoressources au sens large. Les systèmes de modélisation actuels offrent des solutions indispensables pour construire des modèles structuraux grâce à des algorithmes avancés, principalement basés sur des concepts mathématiques.

Cependant, ces systèmes reposent fortement sur les interprétations des experts. Les tâches interprétatives souvent employées tacitement par les experts, sont répétitives, fastidieuses, mais également empreintes de subjectivité. Le problème réside dans le fait que les systèmes utilisés ne capturent pas la connaissance implicite, ce qui les empêche de contribuer à l'interprétation. De plus, ce qui est enregistré dans le modèle final est la description mathématique des représentations des objets modélisés, et la majorité des décisions prises n'est jamais enregistrée. Donc, il est impossible de savoir ce qui a été interprété à chaque étape, avec quel degré de certitude, et sur quelle base d'information. Cette manière de modéliser restreint la possibilité de l'intégration de la connaissance tant dans la structure interne des systèmes que pour les utiliser durant la construction d'un modèle. En effet, seules les connaissances directement formulables dans le formalisme mathématique en place sont souvent implémentées. Une grande partie des connaissances expertes, bien qu'essentielle, ne peut être facilement incorporée, ce qui entrave le partage, la mise à jour et la reproduction des modèles. Les différences de vocabulaires, formats et algorithmes posent également des problèmes d'interopérabilité, rendant difficile la collaboration entre experts.

Ces verrous scientifiques peuvent être résumés en deux axes de recherche : (1) la formalisation du processus d'interprétation et (2) la formalisation de la connaissance des experts. En premier lieu, nous proposons un formalisme pour le processus d'interprétation axé sur une approche itérative et évolutive qui recherche, sélectionne et explique les situations ambiguës ou manquantes d'informations, en s'appuyant sur un modèle de connaissance externe. Ce modèle intègre de manière structurée et explicite les connaissances nécessaires pour décrire les objets géologiques et valider les modèles. Un prototype est développé et appliqué sur des données de terrain pour montrer le potentiel de cette approche à capturer et automatiser diverses tâches interprétatives, ouvrant la voie à une automatisation complète du système.

Le deuxième volet de ce travail de recherche se concentre sur la formalisation des différents aspects de la connaissance des experts identifiés précédemment. Pour ce faire, nous avons recours à l'ingénierie des connaissances, en particulier les ontologies. Un cadre ontologique est proposé combinant une ontologie existante, décrivant les objets géologiques, avec une ontologie développée spécifiquement pour ce travail, nommée POKIMON, pour capturer la connaissance nécessaire à modéliser ces objets. POKIMON se concentre sur l'explicitation des notions des modèles, leurs représentations, les relations entre ces représentations, et les processus de modélisation. Ce cadre facilite ainsi l'automatisation du système d'intégration proposé précédemment.

Ce travail de recherche explore une approche de la modélisation en s'appuyant sur la formalisation et la modélisation conceptuelle. Les résultats obtenus montrent un potentiel intéressant, mais pour en tirer pleinement parti, il est nécessaire d'améliorer plusieurs aspects, tels que la prise de décision et le développement approfondi de la notion des situations interprétables. Également, plusieurs pistes de développement potentielles ont été identifiées, notamment la réalisation d'enquêtes auprès d'experts pour mieux décrire l'aspect stratégique de la prise de décision, ainsi que l'application de méthodes d'apprentissage automatique pour des tâches ciblées dans le cadre de l'algorithme développé.

Abstract

In the field of geology, three-dimensional structural modeling is an essential tool to represent the subsoil architecture, and to improve communication and understanding of experts. Current modeling systems offer a myriad of powerful solutions to build structural models using advanced algorithms, though, these systems heavily rely on expert interpretations.

During interpretation, experts perform several tasks often implicitly, such as selecting and refining inputs, determining the order of interpretation, and identifying geological objects. These tasks are typically subjective and prone to error. Although subjectivity itself is not inherently problematic, the issue lies in the fact that existing 3D geomodelling systems are not equipped to capture the implicit application of knowledge. As a result, they are unable to perform interpretations autonomously or to meaningfully contribute to interpretative tasks. Moreover, the final model only records the mathematical representation of the modeled objects, while most decisions made during the process are not documented. Consequently, it is impossible to know what was interpreted at each step, the degree of certainty involved, or the basis of the information used.

This modeling approach also limits the integration of knowledge both at the backend of systems and during model construction when experts want to share their interpretations. Typically, only knowledge that can be directly formulated in the existing mathematical formalism is often implemented. Accordingly, a large portion of essential knowledge cannot be easily incorporated, hindering the sharing, updating, and reproduction of models. Additionally, differences in vocabulary, formats, and algorithms create interoperability problems, making collaboration between experts more challenging.

These challenges can be categorized into two research areas: (1) the formalization of the interpretation process and (2) the formalization of expert's knowledge. First, we propose a formalism for the interpretation process focused on an iterative and incremental approach that searches for, selects and explains ambiguous or information-lacking situations, based on an external knowledge model. This model integrates in a structured and explicit way the knowledge necessary to describe geological objects and validate models. A prototype was developed and applied on field data, showing the potential of this approach to capture and automate various interpretative tasks.

The second part of this research focuses on formalizing the different aspects of expert knowledge identified earlier. For this, we use knowledge engineering, particularly ontologies. An ontological framework is proposed that combines an existing domain ontology, describing geological objects, with a new ontology developed specifically for this research, named POKIMON. POKIMON captures the knowledge needed to model these geological objects, focusing on explicit description of the concepts, their representations, the relationships between representations, and the associated modeling processes. This framework contributes to the further automation of the previously proposed system.

This research explores an approach to geomodelling based on formalization and conceptual modeling. Obtained results show significant potential, but several aspects need improvement to maximize its benefits, such as decision-making and the detailed development of the concept of situations requiring interpretation. Several potential development paths have been identified, including conducting surveys among experts to enhance the decision-making aspect during interpretation and applying machine learning methods in specific contexts to improve the performance of targeted tasks within a structured framework.

Résumé étendu en Français

1. Introduction

La modélisation est un outil essentiel en géologie, notamment pour la construction de modèles numériques structuraux qui représentent l'architecture du sous-sol. Cette architecture désigne la configuration et l'arrangement des unités géologiques et leurs relations structurales. La modélisation structurale permet aux experts de communiquer leur connaissance géologique et de mieux comprendre la configuration géologique de la zone étudiée à travers différentes méthodes de représentation et de calcul scientifique des propriétés des objets géologiques étudiés. La modélisation des structures géologiques constitue donc le socle des informations utiles aux nombreuses applications géoscientifiques au cœur de différents enjeux économiques et sociétaux, tels que l'analyse qualitative et quantitative des ressources naturelles, la gestion des ressources aquatiques, et la mitigation des risques naturels, pour ne citer que quelques exemples. De ce fait, la modélisation structurale est un sujet scientifique en constante évolution avec l'élaboration de nouvelles méthodologies visant à améliorer la construction, la compréhension et l'utilisation des modèles structuraux.

En pratique, les modèles structuraux sont construits à partir de différents types de données, telles que les relevés de terrain, les forages, les données géophysiques et les cartes géologiques. Le processus de construction de ces modèles peut être simplifié en deux grandes étapes : l'interprétation des données par les experts, suivie de la construction des représentations géométriques de ces interprétations à l'aide d'outils de modélisation. Ces deux étapes ne sont souvent pas exécutées de manière linéaire, mais plutôt de façon itérative durant la construction des modèles. De nombreux outils et environnements numériques facilitent aujourd'hui la création de ces modèles. Les objets et structures géologiques tels que les unités et les failles, sont généralement représentés dans les modèles par des surfaces structurales (e.g., toit, mur, faille, surface extérieur pour les corps intrusifs) en utilisant différentes méthodes de représentation.

Dans leur revue, Wellmann et Caumon (2018) discutent les deux principales approches de représentation des interfaces (surfaces structurales) : l'approche explicite et implicite. L'approche explicite se base sur la représentation des surfaces comme des objets géométriques, en utilisant par exemple des fonctions de profondeur (Hardy, 1971), des surfaces triangulées (Mallet, 1992) ou des surfaces paramétriques (De Kemp, 1999). À l'inverse, l'approche implicite repose sur l'utilisation d'isovaleurs d'un champ scalaire défini par une fonction implicite pour représenter les surfaces (Calcagno et al., 2008; Lajaunie et al., 1997). Les représentations explicites offrent plus de contrôle sur les caractéristiques locales et la topologie (Collon et al., 2016) alors que les méthodes implicites facilitent plutôt la mise à jour des modèles et l'incorporation de plusieurs concepts géologiques comme les paramètres structuraux de plis (Laurent et al., 2016).

Pour générer des valeurs dans les endroits non couverts par des données quelque soit la méthodes de représentations des objets (explicites ou implicites), de nombreuses techniques d'interpolation ont été développées (Franke, 1982; Lajaunie et al., 1997; Lyu et al., 2021; Mallet, 1997; Oliver and Webster, 1990; Sprague and De Kemp, 2005; Watson, 1999). En effet, l'interpolation est un concept fondamental dans la construction des modèles vu la quantité des données initiales.

Il est important de noter que chaque technique d'interpolation a ses propres avantages et inconvénients (Wellmann and Caumon, 2018). Par exemple, le Krigeage permet de prendre en compte les corrélations spatiales et la variabilité entre les données d'entrée à une échelle globale ou locale (Fabbri et al., 2020; Lajaunie et al., 1997; Renard et al., 2019). Par ailleurs, l'utilisation de fonctions de base radiales (RBF) (Franke, 1982; Hillier et al., 2014) met l'accent sur les influences localisées. Cette méthode est particulièrement efficace pour capturer des variations lisses et continues, notamment dans un voisinage local bien défini. Toutefois, lorsque ces techniques sont appliquées en utilisant un voisinage global, elles peuvent produire des résultats similaires en termes de distribution interpolée, bien que leurs hypothèses et structures mathématiques sous-jacentes diffèrent.

Malgré la puissance des outils basés sur ces méthodes et techniques de construction des représentation géométriques, ils manquent intrinsèquement de capacité d'interprétation, nécessitant ainsi une implication substantielle des experts. Des travaux récents, tels que ceux de Laurent et al. (2016) et Pizzella et al. (2022) intègrent des concepts géologiques pour améliorer la caractérisation des structures, en particulier des plis. Cependant, l'identification et la validation des types de déformation sont toujours réalisées par les experts. Certaines solutions visant à formaliser certaines parties du processus d'interprétation ont été développées, mais elles se concentrent sur un seul type de structure, comme la détection de failles (Cherpeau, 2012; Godefroy et al., 2019) ou l'estimation des caractéristiques des plis (Grose et al., 2018, 2017).

Même avec des outils intégrant certains aspects de la connaissance pour aider à interpréter une caractéristique donnée, les résultats – les représentations géométriques– ne sont pas toujours valables géologiquement (De Kemp et al., 2017; Zhong et al., 2020). Par exemple, les interfaces générées par des méthodes explicites nécessitent l'implication d'experts qui utilisent leurs connaissances individuelles pour ajuster la géométrie et la topologie des interfaces (e.g., découpage, lissage) (Caumon et al., 2009) et pour établir des relations entre ces interfaces (e.g., recoupement) (Caumon et al., 2004). D'autre part, les résultats des méthodes implicites nécessitent l'intervention d'experts pour valider les géométries de structures complexes (par exemple, l'ajustement de surfaces fermées) (Belhachmi, 2024).

Les efforts pour intégrer davantage de concepts dans la représentation des structures, tels que ceux de Laurent (2013), Laurent et al. (2016) et Grose et al., (2021b), sont limités par les cadres mathématiques centrés sur l'interpolation ou les défis posés par les approches inverses (De La Varga and Wellmann, 2016; Wellmann and Caumon, 2018). Par exemple, il est difficile de prendre en compte les contraintes sur les épaisseurs dans un système linéaire d'équations sans recourir à un processus itératif (Laurent, 2016). De même, l'utilisation de données plus complexes que des données éparses reste limitée dans les approches inverses (Caumon, 2010).

2. Verrous scientifiques

Bien que les outils de modélisation actuels offrent des solutions incontournables pour construire des modèles structuraux à l'aide d'algorithmes avancés, ils restent incapables de réaliser l'interprétation. Lors de l'interprétation, les experts choisissent quelles données utiliser, quels endroits ou situations nécessitent une explication, et dans quel ordre commencer l'interprétation du modèle. De plus, c'est aux experts d'identifier les objets et structures géologiques en se basant sur leurs expressions partielles dans les données, et de s'assurer de la validité des résultats d'interpolation. Ces tâches sont souvent redondantes, fastidieuses, mais surtout sont aussi empreintes de subjectivité. Cela s'explique par le fait que les experts sont des êtres humains, et que leurs décisions sont basées sur leurs niveaux de connaissance, d'expertise, mais aussi sur leurs processus mentaux et cognitifs, comme discuté par Bond et al. (2007). Ainsi, les résultats sont souvent sujets à des biais cognitifs (Frodeman, 2014).

Bien que la subjectivité puisse parfois être perçue comme une limitation, elle représente également une opportunité, car elle permet aux experts de réfléchir de manière créative et sortir des sentiers battus. Le véritable problème réside dans l'incapacité des outils de modélisation à capturer ces processus mentaux et cognitifs. En conséquence, une grande partie des décisions prises par les experts échappe à toute documentation dans le modèle. Cela complique non seulement la mise à jour et la révision par les pairs, mais également la préservation de l'historique complet des étapes de construction. En effet, les modèles n'enregistrent que le résultat final sous forme de fonctions mathématiques décrivant les surfaces représentées. Cette approche ne permet pas de savoir ce qui a été interprété, avec quel niveau de certitude, ni sur quelle base d'informations. Par conséquent, la méthode actuelle de construction des modèles limite significativement leur utilité dans la phase d'interprétation et réduit leur capacité à assister les experts à un niveau conceptuel plus avancé.

Une autre conséquence de ce paradigme de modélisation est la contrainte de représentation de la connaissance, tant pour l'introduire dans l'outil de modélisation que pour l'intégrer durant le processus de l'interprétation. En effet, seule la connaissance formulable mathématiquement est généralement implémentée dans les outils. Cependant, une grande partie des connaissances, pourtant très basiques dans un processus d'interprétation, demeure détenue exclusivement par les experts, faute de moyen pour l'introduire. Un exemple simple est l'épaisseur des couches qui, selon Laurent (2016) et De La Vargua & Wellmann (2016), est difficile à intégrer dans un formalisme basé sur l'interpolation, nécessitant de lourdes phases d'itération.

Une autre limitation des outils de modélisation associés au paradigme basé sur l'interprétation humaine réside dans les problèmes de partage, de mise à jour, et de reproduction de ces modèles. En effet, les outils de modélisation utilisent souvent des formats d'encodage spécifiques et des méthodes diverses pour décrire les objets et structures géologiques, même lorsqu'ils emploient les mêmes algorithmes de calcul. Cela entraîne de sérieux problèmes d'interopérabilité entre les systèmes, de partage des connaissances, et de collaboration entre experts.

3. Objectifs de la thèse

Ces verrous scientifiques peuvent être résumés en deux axes de recherche : le premier se concentre sur le processus d'interprétation et le deuxième sur la connaissance des experts. L'objectif est d'aborder ces deux aspects du point de vue de la formalisation et la modélisation conceptuelle.

Dans un premier lieu, nous proposons un formalisme pour le processus d'interprétation, en nous inspirant de la façon dont les experts en modélisation interprètent les données de modélisation. L'idée donc est de considérer les problèmes de modélisation d'un point de vue d'interprétation et non pas d'un point de vue de l'encodage mathématique des connaissances experts. Cette approche amène vers une interaction plus profonde entre les systèmes et les interpréteurs, impliquant une compréhension conceptuelle approfondie du processus de modélisation et une meilleure intégration des connaissances. Ces éléments sont essentiels pour automatiser le processus d'interprétation et améliorer le partage des connaissances entre experts et systèmes.

Adopter cette perspective interprétative implique de formaliser le processus d'interprétation comme une entité distincte, en se basant sur des connaissances structurées qui encapsulent la connaissance des experts. Cette thèse établit un cadre de connaissances simple, axé sur un modèle générique décrivant comment les experts en modélisation structurale 3D s'engagent dans l'interprétation. Ce cadre inclut les connaissances géologiques, interprétatives et représentationnelles, détaillées plus en profondeur dans la thèse.

L'objectif aussi de cette formalisation du processus d'interprétation est de mettre en lumière les 'situations interprétables' et appelle à une analyse conceptuelle plus profonde des relations entre les modèles, les représentations et les processus impliqués. Pour formaliser ces différents aspects, nous avons recours à l'ingénierie de la connaissance, en se basant sur les ontologies (Guarino and Giaretta, 1995). Cette approche, bien qu'émergente dans les géosciences (Abel et al., 2012b), nécessite une modélisation conceptuelle rigoureuse. Il est donc crucial de développer des modèles de connaissances fiables, alignés sur les exigences spécifiques du domaine de la modélisation 3D et de la géologie, tout en assurant leur réutilisabilité et interopérabilité avec d'autres applications géoscientifiques.

4. Formalisme d'interprétation

4.1. Théorie

Le formalisme d'interprétation proposé s'inspire du processus d'interprétation tel qu'il est réalisé par les experts. Les points clés sont résumés ci-après.

- Tout d'abord, ce que les experts interprètent, ce sont des situations inexpliquées dans le contexte de la modélisation et qui découlent par exemple de : la dispersion et manque de données, discontinuités des objets géologiques, par exemple deux parties stratigraphiques isolées de la même unité géologique ; des objets inconnus ou des propriétés manquantes d'entités géologiques, par exemple une partie stratigraphique sans une attribution à une unité stratigraphique; ou une ambiguïté dans l'expression des objets géologiques dans les données, par exemple une couche sédimentaire inclinée sans que cela soit expliqué par un type de déformation. Ces situations ne répondent pas à certains principes ou hypothèses géologiques de base, par exemple le principe d'horizontalité d'origine ; ou ne fournissent tout simplement pas suffisamment d'informations dans le contexte de la modélisation. L'interprétation de ces situations implique de proposer certaines entités (objets, structures ou propriétés) ou de mettre à jour les propriétés d'entités existantes pour fournir une explication dans le contexte de la modélisation. Par exemple, la situation constituée de deux parties stratigraphiques isolées peut être expliquée par une partie plus grande qui couvre l'écart et qui devrait être cohérente avec les deux parties stratigraphiques initiales. D'autre part, la couche sédimentaire inclinée peut avoir une explication possible par l'existence d'une structure de déformation (par exemple, un pli, une faille). Ce formalisme repose donc sur l'existence d'un nombre fini de possibilités conceptuelles de situations expliquant les concepts et qui sont identifiés et instanciés au moment de l'interprétation. Par exemple, des couches incurvées sont expliquées par une possible instance de plis, mais le concept de plis n'est pas inventé à chaque fois.

- Pour interpréter ces situations les experts sélectionnent une situation donnée, l'analysent, identifient les ambiguïtés possibles et formulent les hypothèses potentielles qui pourraient expliquer cette situation. Ensuite, ils choisissent une théorie parmi les options disponibles, l'évaluent, puis proposent cette théorie comme explication. Trois étapes principales émergent : (1) la planification du processus, (2) la réalisation du processus et (3) l'évaluation de la réalisation. Les experts utilisent les nouvelles informations issues d'une itération d'interprétation pour mettre à jour leur compréhension (par exemple, lorsqu'une nouvelle structure est identifiée). Par conséquent, une quatrième étape peut être envisagée, où les experts réévaluent leur compréhension et décident de l'orientation de leur processus d'interprétation. En anglais, ces quatre étapes sont désignées par l'acronyme PDCA (Plan, Do, Check, Act) en référence à la théorie des processus d'amélioration continue développée par Deming (Deming, 1997; Moen, 2009).
- Le processus d'interprétation est itératif et incrémental, permettant aux experts d'extraire de nouvelles informations et d'améliorer continuellement leur compréhension. Ils utilisent également leur mémoire des itérations précédentes pour éviter de répéter les mêmes erreurs.
- L'interprétation est un processus continu, où les experts révisent leurs hypothèses lorsque de nouvelles incohérences sont découvertes à travers d'autres itérations d'interprétation. Cela montre que les interprétations sont souvent temporaires et sujettes à modification.
- Les experts formulent d'abord des interprétations théoriques, qu'ils adaptent ensuite pour les aligner aux contraintes des outils de modélisation. Ce processus met en lumière l'importance de dissocier l'interprétation conceptuelle de sa représentation géométrique avec un outil quelconque.

En résumé, le formalisme proposé considère l'interprétation comme un processus itératif et incrémental, avec des résultats ouverts à la révision. Un retour d'information est essentiel, tant sur les objets interprétés que sur le processus d'interprétation lui-même.

Pour automatiser ce processus, un système numérique flexible et un modèle de connaissances contextuel sont nécessaires pour guider les interprétations. Ce modèle de connaissance devrait décrire trois aspects majeurs : ceux liés à la géologie, à l'interprétation et à la représentation. Les aspects géologiques doivent décrire la réalité géologique, y compris les objets, structures et leurs caractéristiques (comme une couche stratigraphique ou un pli). Cet aspect concerne les éléments à modéliser dans le cadre de l'interprétation géologique. Les aspects d'interprétation doivent expliquer comment les situations géologiques sont interprétées. Cela inclut la description des situations nécessitant une explication, les concepts explicatifs associés et la traçabilité de l'histoire des interprétations pour une future référence. Les aspects de représentation doivent fournir un cadre spatial et temporel pour explorer les aspects géologiques et interprétatifs, par exemple, en représentant les objets dans un système de coordonnées spécifique ou en visualisant les interprétations à chaque étape.

4.2. Implémentation et application

Durant ce travail de recherche, nous avons développé un algorithme qui implémente les aspects discutés auparavant. Similairement à l'idée de la séparation du modèle de sa présentation, une idée principale qui guide l'implémentation de cet algorithme est la séparation des taches de leur façon de faire. Par exemple, sélectionner des données à interpréter est une tache interprétative dont l'exécution peut être implémentée selon plusieurs stratégies. L'algorithme proposé est composé de trois blocs principaux : un gestionnaire de connaissance géologique, un gestionnaire d'interprétation et un gestionnaire de représentation et de visualisation spatiale.

- Le gestionnaire de connaissance gère l'interaction avec une base de connaissance extérieure sous forme d'une ontologie. Au stade initial, il convertit les données des bases de données externes vers de nouvelles instances de l'ontologie sous-jacente. Ce gestionnaire fournit des fonctionnalités pour construire, mettre à jour et gérer les caractéristiques géologiques en fonction de leurs définitions dans le modèle de connaissance.
- Le gestionnaire d'interprétation applique le processus d'interprétation formalisé ci-dessus. Chacun de ces sous-processus –PDCA– s'appuie sur des fonctionnalités fournies par le gestionnaire de connaissance (par exemple, les constructeurs et les éditeurs).
- Le gestionnaire de représentation et de visualisation fournit des espaces pour la représentation. Une information peut être associée à plusieurs espaces tandis que chaque espace possède ses propres paramètres (par exemple, système de coordonnées, résolution) avec conversion possible entre eux. Les propriétés de l'espace peuvent être dérivées des entrées ou définies par les utilisateurs. Ce gestionnaire est chargé de fournir une visualisation numérique du modèle et des informations encapsulées, qui peuvent être en une à trois dimensions.

Pour l'application de cet algorithme, nous avons utilisé une ontologie simplifié nommée MOGI (Minimal Ontology for Geological Intepretation). Cette ontologie encapsule des aspects de connaissance liés à un cas d'application sur des plis de type chevron.

Dans les résultats de notre approche, le concept de "plis en chevron" défini dans le modèle de connaissances guide et justifie le style du plissement. Les propriétés conceptuelles des structures (surfaces axiales, axes de plis, membres) utilisées par le système pendant l'interprétation font partie du modèle et peuvent être interrogées. De plus, notre approche aborde le problème du manque de données et de la continuité possible des objets en définissant explicitement les concepts et les règles qui gouvernent leurs interactions. Cela améliore le contrôle des résultats dans les régions à forte incertitude épistémique. En appliquant pleinement le formalisme d'interprétation proposé, toutes les connaissances utilisées dans chaque itération sont stockées dans la base de connaissances. Ces informations peuvent être consultées pour récupérer des détails sur le cycle d'interprétation à l'origine de chaque partie du modèle, ainsi que sur les théories disponibles et celles choisies.

En substance, l'approche proposée améliore considérablement les systèmes de modélisation géologique existants car elle est (1) sémantiquement riche et cohérente, (2) exploite les concepts géologiques pour affiner la géométrie construite, et (3) explicite le processus décisionnel ayant guidé sa construction.

Bien que ce premier axe de recherche ait été consacré à la production d'un formalisme générique pour le processus d'interprétation, le fait que ce processus soit intrinsèquement lié à la connaissance experte a permis d'établir un état des lieux préalable de cette connaissance, et de définir ce qu'il est nécessaire de formaliser et de développer dans le cadre de l'ingénierie des connaissances, en particulier :

- Quelles connaissances sont essentielles pour construire des modèles 3D et leurs représentations ?
- Comment modéliser ces connaissances à l'aide de langages d'ontologie formels ?
- Comment les processus employés pour construire des modèles peuvent-ils être représentés à l'aide de langages formels ?

À partir de ces questions fondamentales, les exigences suivantes sont dérivées pour construire un modèle ontologique qui y répondra, tout en fournissant l'infrastructure de connaissances nécessaire au processus de modélisation, incluant les aspects liés à l'interprétation :

- Un cadre qui structure les connaissances sur les entités géologiques, les observations et les situations à interpréter.
- Un cadre qui décrit la construction d'une interprétation et son lien avec les connaissances géologiques ainsi qu'avec les entités existantes ou non dans la réalité.
- Un cadre pour gérer les entités soumises à l'évaluation et à la décision concernant les hypothèses d'interprétation et les différentes parties du modèle.
- Un cadre pour décrire les différents types de modèles en fonction de leur aptitude à être représentés et les relations entre ces représentations.

- Un cadre pour rendre compte des différents processus responsables de la génération et de la gestion de tous ces aspects de modélisation. Ce cadre doit permettre de décrire ces processus, leur fonctionnement et leurs spécifications.

Avant de présenter les résultats de notre travail pour répondre à ces aspects, la section suivante présentera un bref aperçu de l'ingénierie des connaissances, des ontologies, ainsi que des principales avancées dans leurs applications pour la modélisation 3D.

5. Ingénierie de la connaissance et formalisation des connaissances

L'ingénierie des connaissances est un domaine de recherche en pleine évolution dans le domaine de l'intelligence artificielle (IA), en particulier l'IA symbolique, et récemment dans l'explicabilité et l'amélioration des technologies d'IA par apprentissage. L'objectif principal de l'ingénierie des connaissances est de concevoir et de construire des systèmes basés sur la connaissance qui peuvent utiliser celle-ci pour traiter les données de manière similaire à ce que ferait un humain. Ce domaine implique donc l'acquisition des connaissances d'experts humains, leur formalisation et leur encodage à l'aide de langages de représentation appropriés, ainsi que leur intégration dans des systèmes (Studer et al., 1998). Dans cette section, l'accent est mis sur la formalisation à l'aide d'ontologies, car elle constitue l'essentiel du travail sur la connaissance dans cette recherche.

5.1. Les ontologies

Les ontologies sont des formalismes essentiels en sciences de l'information pour structurer et formaliser les connaissances. Studer et al., (1998) décrivent l'ontologie comme une "spécification formelle et explicite d'une conceptualisation partagée". Les ontologies permettent de capturer et de structurer la connaissance d'un domaine spécifique, offrant une description standardisée des concepts, de leurs relations, et des règles de fonctionnement. Elles réduisent l'ambiguïté, facilitent l'interopérabilité entre systèmes, permettent l'intégration et la réutilisation des données, et offrent des capacités d'interrogation et de raisonnement. Les ontologies sont structurées autour de cinq types de composantes clés :

- **Classes** : Représentent les concepts dans l'ontologie (par exemple, des concepts géologiques). Les classes sont définies en fonction du niveau d'abstraction, de l'atomicité et du niveau de réalité, la hiérarchisation et également variabilité.

- **Relations** : Définissent les interactions entre les classes, et par conséquence entre les instances de ces classes. Elles peuvent être taxonomiques (hiérarchiques) ou associatives (non hiérarchiques), et incluent des propriétés comme la symétrie et la transitivité.

- **Règles** : Énoncés qui définissent des conditions supplémentaires pour imposer des contraintes logiques, souvent exprimées avec des langages formels comme le SWRL (Semantic Web Rule Language).

- **Axiome** : Assertions constantes sur les entités de l'ontologie, servant à définir des règles d'inférence et des contraintes de cardinalité, telles que le nombre d'instances possibles d'une propriété.

- **Instances** : Représentent des individus spécifiques au sein des classes définies dans l'ontologie.

L'infrastructure numérique pour implémenter, manipuler et gérer les ontologies est assurée par les technologies de web sémantiques, en particulier, les IRI, RDF, OWL, et SPARQL (Breitman et al., 2007).

5.2. La formalisation des connaissances pour la modélisation 3D

L'utilisation des connaissances dites formelles en géosciences est relativement récente par rapport aux autres domaines tel que la médecine et la biologie. Les travaux existants sont souvent conduits par la nécessité du partage, le stockage et le requêtage des données géologiques. Le standard RESQML (Morandini et al., 2011) constitue une contribution majeure au partage de données dans le domaine de la modélisation, en particulier dans le secteur pétrolier. Toutefois, il ne peut être considéré comme une véritable ontologie en raison de l'absence d'éléments liés à la modélisation conceptuelle et au raisonnement.

Les ontologies sont de plus en plus employées car elles offrent un cadre sémantique robuste. Des travaux comme ceux de Zhan et al. (2022, 2019), se concentrent sur la formalisation des propriétés mathématiques des modèles 3D pour capturer la sémantique géologique et géométrique afin de l'aligner à des ontologies existantes telle que celle de Babaei et al. (2006). D'autres travaux, comme ceux de Cicconeto et al. (2022) et Qu et al. (2023), développent des ontologies spécialisées pour la modélisation des réservoirs. D'autres travaux des ontologies ont été développés pour d'autres applications dans le domaine des géosciences et démontrent des résultats prometteurs (Abel et al., 2012a, 2019; Babaie et al., 2020; Belaid et al., 2009; Fauziati and Watanabe, 2010; Lombardo et al., 2018; Ma et al., 2012; Mantovani et al., 2020; Mastella et al., 2009; Simons et al., 2006; C. Wang et al., 2018; Zhong et al., 2009).

En règle générale, les ontologies en géosciences sont souvent conçues pour des applications spécifiques et beaucoup restent isolées malgré qu'elles décrivent souvent des entités géologiques communes entre les différentes applications. Cela est dû à l'absence historique d'ontologies du domaine des géosciences. Pour combler ce manque, deux ontologies du domaine des géosciences ont été développées : GSO (Brodaric and Richard, 2020) et GeoCore (Garcia et al., 2020). Ces ontologies fournissent une base solide pour la création d'applications plus spécifiques, mais leur adoption reste limitée.

L'incertitude et l'interprétation, intrinsèquement liées à l'information géologique, constituent des défis majeurs, notamment en modélisation structurale 3D. Par conséquence, l'utilisation des ontologies en géosciences doit aller au-delà de simples étiquettes catégorielles pour inclure les aspects épistémologiques et interprétatifs des informations géologiques. Les travaux existants négligent souvent l'interprétation et les processus de modélisation qui sous-tendent les représentations numériques. Seuls quelques travaux, tels que ceux de Brodaric (2012) et Carbonera et al. (2013), abordent explicitement ces questions. Ces études abordent des aspects de l'interprétation et de l'inférence dans la construction des connaissances géologiques, mais la modélisation de l'interprétation en tant qu'entité distincte et la gestion de l'incertitude dans les informations géoscientifiques demeurent des domaines encore peu explorés.

6. POKIMON une ontologie pour formaliser la connaissance en modélisation 3D

POKIMON (pour Processes, Observations, Knowledge, Information, Modeling, ONtology) est une ontologie développée spécifiquement pour structurer la connaissance experte utilisée lors de la construction de modèles 3D géologiques. Elle s'intéresse particulièrement aux aspects liés à la définition des modèles géologiques structuraux, à leurs caractéristiques conceptuelles et spatio-temporelles, ainsi qu'à l'impact de ces caractéristiques sur leur représentativité. Elle fournit également un cadre pour structurer les processus utilisés durant l'interprétation, avec un accent sur les processus interprétatifs.

Concrètement, cette ontologie a pour but de modéliser les aspects clés suivants et les entités potentielles :

- Modèle conceptuel : modèle géologique avec des propriétés spatiales (et éventuellement temporelles) incomplètes, comme un log stratigraphique.
- Modèle spatio-temporel : doté de propriétés spatiales et temporelles complètes.
- Omission : exclusion d'objets dans l'interprétation qui ne respectent pas certaines conditions.
- Simplification : remplacement d'objets complexes pour but de représentation avec des entités non existantes en réalité.
- Intégrité du modèle après modification et quand il est considéré comme un autre modèle différent.
- Situation à interpréter et sa relation avec les objets réels, et les lois géologiques.
- Le processus d'interprétation et son historique, ses caractéristiques et relation avec les autres processus.
- Évaluation de la validité d'un modèle et lui associer une décision en fonction de l'évaluation.
- L'interpolation comme processus d'estimation et de la construction de géométries.

6.1. Méthodologie

POKIMON est fondée sur des ontologies de haut niveau (top-level) comme la Basic Formel Ontologie (BFO) (Otte et al., 2022) et de niveau intermédiaire (mid-level) comme Information Artifcat Ontologie (IAO) (Ceusters and Smith, 2015) et Industry Ontology Foundary (IOF) (Drobnjakovic et al., 2022), elles-mêmes basées sur BFO. Ces ontologies se concentrent sur la représentativité de l'information, ce qui est en lien direct avec l'objectif de représenter et modéliser la connaissance autour des modèles 3D. Pour décrire les entités géologiques, POKIMON s'appuie sur la GeoSience Ontology (GSO) (Brodaric and Richard, 2020), une ontologie dédiée à la description des objets, structures et relations géologiques. Les concepts pertinents sont extraits et étendus pour garantir l'indépendance de cette ontologie et sa cohérence interne, car les ontologies existantes sont parfois mal alignées et évoquent des aspects éloignés de notre application.

Entités principales dérivées de BFO, IAO et IOF :

- Information Content Entity (ICE) : Entité de contenu d'information, représentant des informations abstraites (e.g., un roman).
- Information Quality Entity (IQE) : Entité de qualité d'information, qualifiant une ICE spécifique, par example, le motif de l'encre sur une page d'un livre.
- Information Artefact (IA) : Artefact d'information, support physique ou numérique contenant des ICE (e.g, la partie physique d'un livre).
- Process : quelque chose qui fait un changement sur le monde.
- Quality : une prospérité d'un objet, elle spécifiquement dépendante à cet objet.

Entités principales dérivées de GSO :

- Geological Unit : Unité géologique, une classe d'entités représentant des formations géologiques.
- Geological Structure : Structure géologique, une classe d'entités représentant les structures géologiques tels que les plis et les failles.

6.2. Développement et résultats

POKIMON distingue fondamentalement les objets et relations dans la réalité de leurs analogues interprétés dans une représentation. Une entité représentative est liée à une réalité (*is about*). Les entités dans une représentation sont divisées en composants abstraits et figuratifs, car une entité interprétée peut être représentée de différentes manières. À partir de cette distinction, POKIMON utilise ces concepts clés :

- ***Geomodel*** : Une *Information Content Entity (ICE)* qui représente des entités géologiques réelles. Il se compose d'un *Semantic Component* et d'un *Depictional Component*.

- ***Semantic Content Entity (SCE)*** : Une *ICE* fournissant une description abstraite des entités géologiques. Elle est liée aux entités géologiques via la relation *has Semantic Content*.

- ***Depiction Content Entity (DCE)*** : Une *ICE* qui décrit le contenu représentatif du *geomodel*.

- ***Conceptual Geomodel*** : Définit les concepts géologiques et leurs relations structurales sans détails spatiotemporels complets. Il capture la topologie et le temps relatif mais manque de précision géométrique et de localisation.

- ***Space-Time Geomodel*** : Représente une configuration géologique spécifique avec des propriétés spatiales et temporelles précises. Chaque *Space-Time Geomodel satisfies* un *Conceptual Geomodel*.

Afin de décrire ces modèles géologiques et leurs positionneent par rapport aux règles géologiques et les contraintes imposées durant leurs constructions, POKIMON utilise les concepts suivants :

- ***Information Entity*** : Un *abstract Continuant* qui existe indépendamment de toute représentation, comme les lois naturelles ou les théories scientifiques.
- ***Norm*** : Une sous-classe d'une *Information Entity* qui régit les comportements, y compris les lois géologiques, les principes et les théories.
- ***Constraint*** : Une sous-classe d'une *Information Entity* définissant des critères pour les *geomodels*, principalement en fixant des limites de qualité spatiale (par exemple, longueur, épaisseur).

Ces modèles géologiques sont construits à partir des situations qui émergent à partir des informations disponibles dans un processus itératif (e.g., observation, interprétation). Pour modéliser cela, POKIMON définit :

- ***Situation*** : Un *Independent Continuant* qui inclut des *Semantic Content Entities (SCEs)* comme parties. Il représente un contexte géologique où certaines propriétés sont incomplètes ou dévient des *norms*, nécessitant une interprétation pour hypothétiser ou modifier des entités géologiques afin d'assurer la cohérence.
- ***Origin*** : Une *quality* dans *POKIMON* qui qualifie les entités géologiques selon leur niveau de certitude en trois types : *observed, interpreted*.

Afin de relier et mieux caractériser ces concepts POKIMON utilise principalement les relations suivantes :

- ***Substitutes*** : Une relation où un *geomodel* remplace un autre en raison de limitations de résolution ou de représentation.
- ***Simplifies/Details*** : Relations inverses entre des *geomodels* à différentes échelles de résolution pour des objectifs variés.
- ***Satisfies*** : Une relation où un *geomodel* est conforme à une *information entity*.
- ***Contradicts*** : Une relation où un *geomodel* n'est pas conforme à une *information entity* ou un *geomodel* plus large ; elle est aussi utilisée lorsqu'un *Space-Time Geomodel* ne correspond pas à un *Conceptual Geomodel*.
- ***has Semantic Component*** : Lie un *geomodel* à son *SCE* et connecte le *SCE* aux entités de l'ontologie du domaine géologique.
- ***has Depiction Component*** : Lie un *geomodel* à son *DCE*, et le *DCE* à divers formats de représentation.
- ***is About*** : Une entité initialement définie dans *BFO*, utilisée pour relier conceptuellement un *model* aux entités géologiques réelles, bien qu'elle reste abstraite et non implémentable. Son domaine est étendu pour inclure la nouvelle classe *Information Entity*.

Afin de modéliser les processus POKIMON utilise le concept '*Planned Process* de l'IOF directement pour décrire les processus planifiés par un operateur et qu'ont un objectif fixe et un certain plan d'action à suivre. POKIMON considère tous les processus utilisés dans la construction des modèles 3D comme des processus planifiés, qui peuvent être décrit par un *Algorithm*, une sous classe de *Planned Process*. POKIMON identifie les principaux processus suivants : sélection des entités, proposition de théories, évaluation des modèles, prise de décisions, création et modification des entités, ainsi que le processus global d'interprétation, dont ces étapes sont des composantes fondamentales.

Afin de permettre l'exécution de ces opérations de différentes manières, comme mentionné dans la première partie de ce travail, POKIMON permet qu'un processus planifié partage son objectif avec d'autres processus planifiés, tout en conservant pour chacun un plan d'action spécifique et unique. POKIMON a été appliqué à divers cas pour évaluer son utilité dans la modélisation et l'interprétation 3D. Les résultats sont prometteurs et offrent un cadre structuré pour améliorer la compréhension et la construction des modèles géologiques.

7. Conclusion

Ce travail de recherche explore une approche innovante de la modélisation, basée sur la formalisation du processus d'interprétation géologique et la modélisation des connaissances des experts. Cette recherche vise à répondre aux verrous scientifiques liés aux approches classiques de la modélisation structurale, qui reposent sur l'interpolation et l'interprétation humaine. Bien que ces approches soient efficaces, elles présentent plusieurs limites, par exemple, les systèmes d'interprétation ne peuvent pas interagir avec le processus d'interprétation tel qu'il est réalisé par les experts. En effet, les experts utilisent souvent leurs connaissances de manière tacite, ce qui signifie qu'une grande partie des décisions n'est jamais enregistrée. Cela complique la compréhension de l'ensemble du raisonnement utilisé et empêche la reproduction des modèles. De plus, cette manière de construction des modèles limite le partage des connaissances, tant entre les experts et les systèmes qu'entre les experts eux-mêmes.

Pour remédier à ces limitations, notre recherche propose une approche en deux volets. Tout d'abord, nous suggérons de formaliser le processus d'interprétation par une méthode itérative et incrémentale, qui identifie, sélectionne et explicite des situations ambiguës ou manquante d'information en s'appuyant sur modèle de connaissances externe. Ce modèle permet une représentation structurée et explicite des connaissances nécessaires pour décrire les objets géologiques et valider les modèles. Deuxièmement, nous nous concentrons sur la formalisation des connaissances des experts, particulièrement, les aspects tacites en utilisant un cadre ontologique. Ce cadre combine des ontologies de domaine déjà existantes formant une base pour une nouvelle ontologie, POKIMON, qui capture les connaissances requises pour modéliser les objets géologiques et les différents processus de modélisation associés.

En intégrant ces approches, nous ouvrons la voie à des systèmes de géomodélisation plus automatisés, avec des modèles plus riches sémantiquement. Ces avancées renforcent la capacité de partager, mettre à jour et reproduire les modèles entre experts et plateformes tout en reposant sur un socle de connaissance explicite. Plusieurs pistes d'amélioration ont été identifiées pour améliorer l'approche proposée qui sont discuté à la fin du manuscrit.

Table of content

Table of Figures

Introduction

This research explores an innovative three-dimensional structural modeling paradigm grounded in a structured interpretation framework and the utilization of explicit formalized knowledge.

In geology, 'models' is a concept used for referring to the representations of geological phenomena. They serve as crucial tools in testing and communicating hypotheses, developing quantitative analyses, and enhancing expert's knowledge. Among these, three-dimensional (3D) structural models are particularly important for digitally representing the geometric, topological, and structural characteristics of subsurface geological entities in a static state.

Given the inherent complexity of subsurface architecture, creating an accurate digital 3D structural model requires more than a straightforward translation of input data into a digital format. It necessitates extracting meaningful insights from incomplete or ambiguous information. This process depends on the cognitive expertise of geologists, complemented by specialized geomodelling systems designed to support this task.

Overview

Modeling systems

3D modeling systems are digital environments that provide experts with the necessary framework to convert their understanding to a digital representation that could be visualized, manipulated and edited (Caumon et al., 2004; Mallet, 1992; Verschuren and Butler, 2016). Modeling systems are conceived on the notion of the numerical representation of geological entities; thus, they provide a mathematical framework to accurately construct these representations based on the geometries (e.g., the position and the shape) and the topologies (e.g., the number, the nature and the connectivity) of geological entities following a Boundary-representation formalism (Caumon, 2003; Mallet, 1997). As the subsurface information is inherently incomplete (Bond et al., 2015; Sprague and De Kemp, 2005), modeling tools are designed to estimate the geometric properties of geological entities, for example, using interpolation (Franke, 1982). Modern 3D modeling systems have significantly evolved, offering advanced algorithms that integrate various geometric and topological aspects during the interpolation process to better represent the geological entities (Mei, 2014; Perrin and Rainaud, 2013; Turner and Gable, 2007; Wellmann and Caumon, 2018).

Modeling process

Constructing structural models using modeling systems is a complex process where multiple experts collaborate to understand and represent at best the studied part of the Earth. This process could be classified generically into two main stages (Caumon, 2003; Lidal et al., 2012; Mallet, 2002; Perrin and Rainaud, 2013; Wellmann and Caumon, 2018), as sketched in Figure 1:

- Interpretation: in this stage experts analyze and interpret existing information derived from modeling inputs (e.g., field observations, well logs, cross-sections), to identify, characterize, and describe geological entities and their properties. Experts must continually ask: What aspects are not understood? What are the possible explanations? Are the explanations meaningful within the given context? How to evaluate the plausibility of the explanations? What alternative could exist? Through this iterative questioning, they refine their understanding of geological entities
- Digital model construction: based on interpretation outcomes and existing inputs, experts choose and run suitable modeling algorithms. Expert's understanding is progressively transferred to the numerical version of the model through the parametrization of the algorithm and the update of the 3D rendering.

The two modeling stages are not executed in a linear sequence; rather, they involve a trial-and-error process where adjustments are made until the model is fully constructed.

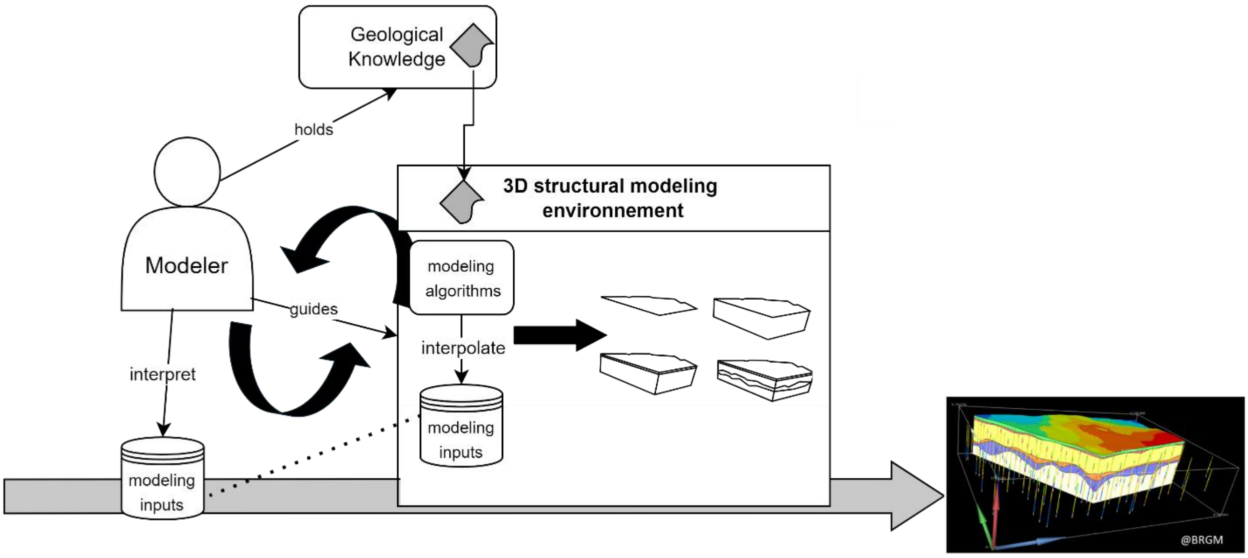


Figure 1 Schematic diagram showing the overall geomodelling workflow via a 3D structural modeling system (environment) based on interpolation algorithms (source 3D model image: BRGM*)

Problem statement

A major challenge that is yet facing the structural modeling process and modeling systems, is the limitation in representing and integrating geological knowledge, whether it be used to guide the interpretation process or to introduce new modeling methodologies and techniques into the system itself. Despite the significant advances in integrating various geological concepts, typically, geological entities that have spatial properties are often targeted, formalized and implemented in modeling systems as these systems are centered on the notion of the 3D representation. However, a substantial portion of essential knowledge –often fundamental to interpretation– remains exclusively held by experts at a conceptual level, due to the absence of mechanisms to incorporate or document it. For instance, thickness is a property that is intuitively used by interpreters to infer the geometries of units, yet integrating it into a linear interpolation framework poses challenges and often requires multiple iterations to better account for it (Laurent, 2013; De La Varga and Wellmann, 2016. This limitation in handling conceptual knowledge is primarily due to the difficulty in developing mathematical formalism to represent such abstract geological features. As a result, modeling systems remain heavily dependent on expert input who interpret, identify, and propose key geological entities and relations (e.g., existing faults, defining the type of the stratigraphic relation), particularly in regions with sparse or incomplete data coverage (Bond et al., 2011) .

This reliance on expert's inputs has numerous drawbacks as, being human, interpreters can make mistakes, and their interpretations and decisions are often subjective. The subjectivity of experts is not inherently problematic (Bond et al., 2011; Frodeman, 1995); the issue lies in the fact that current systems are not designed to account for this subjectivity as they are designed to manipulate outcomes, not to engage with the interpretation process itself. Consequently, these systems cannot understand why interpreters make certain choices or which information have been used to generate a given interpretation, leading to a significant loss of the knowledge employed by experts (Laouici et al., 2024). This further encourages experts to apply their knowledge tacitly, without systematic integration into the model. Even when offering assistance for the interpretation process, these systems often target a specific structure (e.g., folds) with customized algorithms, that are impractical for other geological scenarios or different modeling conditions. Therefore, this paradigm of building models greatly limits the contribution of modeling tools in interpretation and the possibility of assisting experts on a more conceptual level. The lack of assistance becomes a liability when dealing with large datasets or when a model needs to be reproduced or updated as the interpretation process is often cumbersome.

Another important limitation of the modeling systems associated with the paradigm based on human interpretation lies in the problems of sharing, updating, and reproducing these models. Most modeling systems use specific encoding formats and different methods to describe common geological objects and structures, even when they use the same computational algorithms. These limitations are not only from the engineering perspective, but they also impede the advancement of geoscientific understanding by making it difficult for researchers to collaborate and validate findings across different platforms.

These limitations and drawbacks of geological modeling at the current state of science could be resumed into two interconnected main challenges: first, defining how experts interpret and use their knowledge to build their understanding, second, identifying which knowledge experts use and how it could be explicitly introduced to a system.

Proposed approach

To address these scientific challenges, this research proposes a paradigm shift in how modeling systems are structured by integrating conceptual modeling and formalization, as illustrated in Figure 2. The aim is to enhance the identification and structuring of critical expert's knowledge for the 3D structural geomodelling process. A key step to achieve this is the development of an interpretation framework that not only captures but also applies structured knowledge to geological interpretation. Thus, I first propose formalizing the interpretation process to establish a foundation and identify critical aspects of expert's knowledge that are afterward formalized. This approach ensures that both the interpretation process and the expert knowledge are systematically organized and integrated into the modeling workflow.

The first topic is addressed through the development of a formalism that frames and structures the interpretation process based on how 3D modeling experts handle it. The idea is to consider modeling challenges from an interpretation perspective and how understanding is progressively built, instead of solely focusing on enhancing the interpolation framework.

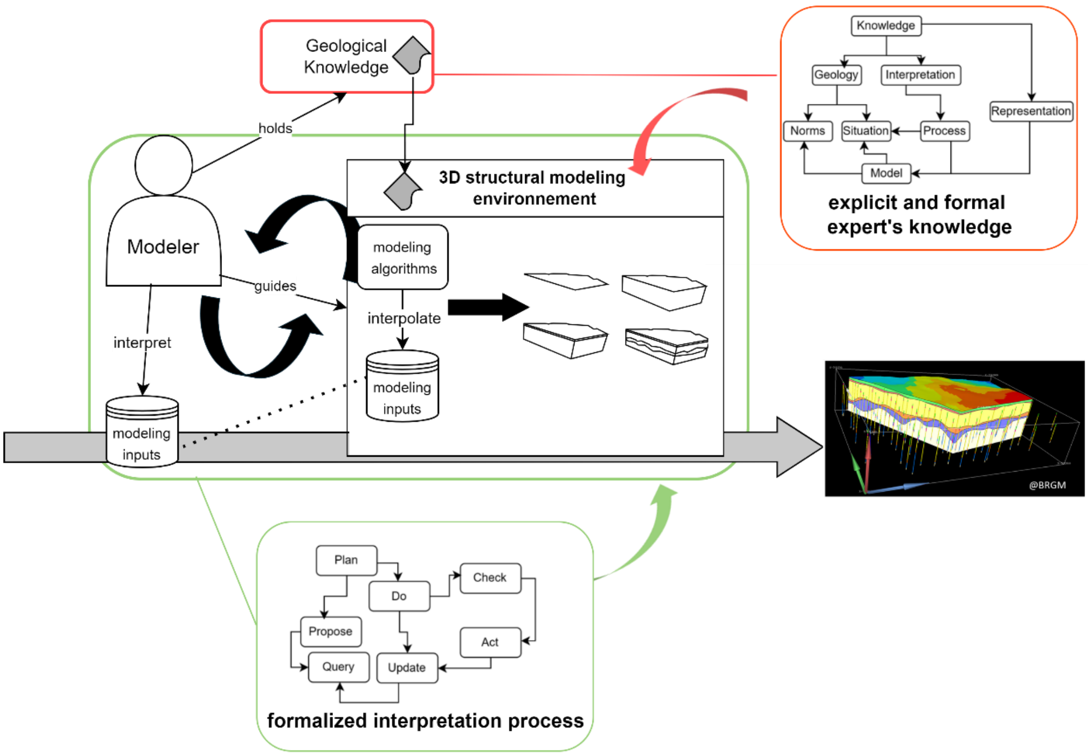


Figure 2 Schematic diagram illustrating the key concepts proposed in this research and their relation to the current 3D modeling workflow. The red block represents the formalization of expert knowledge, while the green block outlines the formalization of the interpretation process with focus on main interpretative tasks that experts undertake

Adopting this interpretative perspective implies formalizing the interpretation process as a distinct entity that refers to an external knowledge model. To achieve this, I develop in this research a system to record the different interpretative tasks, allowing their implementation in different ways, and the ability to execute them automatically and independently from expert's involvement, while also supporting the interaction with experts. To support this process, I establish a simple knowledge model, that explicitly structures geological, interpretative and representational knowledge aspects needed in the interpretation process. This approach fosters a more integrated interaction between systems and interpreters, promoting a deeper conceptual understanding of the modeling process and enhancing the integration of knowledge.

The second topic involves the formal development of identified knowledge aspects within a more robust and structured knowledge framework using ontologies –a semantic technology for formalizing knowledge. This framework is designed to support the interpretation process by providing the necessary knowledge for performing the different geomodelling processes. The framework is structured around the conceptual modeling of expert's understanding, of the geological phenomena, the various aspects related to the notion of models and representations, and the characteristics by which 3D structural models are evaluated and validated. By formalizing this knowledge, the framework aims to capture both explicit and tacit expert knowledge, enhancing the transparency of the geomodelling process.

Communications and Work Dissemination

During this research work, numerous scientific communications were conducted, including conference presentations, journal publications, and collaborative meetings:

Articles:

- A knowledge-driven modeling formalism for automatic structural interpretation, published in Earth Science Informatics, under doi.org/10.1007/s12145-024-01613-y.
- POKIMON: an ontology for geological interpretation in 3D geo-modelling (in preparation for submission to one of these journals: Computers & Geosciences, Geoscientific Model Development, or Applied Ontology.

Conferences:

- IAMG 2022: Oral presentation (no proceedings)
- 6th European 3D Modeling Meeting: Oral presentation (no proceedings)
- FOIS 2024: Oral presentation, poster, and conference paper (with proceedings)

Additional Communications:

- Presentations at the Geological Survey of Canada and the Laboratory of Semantic Technologies of the University of Toronto.
- Two presentations at the Science and Humanities Laboratory, University of Tours, at MADICS meetings.
- Various presentations at hosting institutions for diverse audiences (e.g., New hires Day, local symposiums).

Context of the PhD Research

This research work is the result of a scientific collaboration between two institutions: the French Geological Survey (BRGM) and the Institute of Earth Sciences of Orléans (ISTO). The BRGM is the leading public institution dedicated to applying Earth sciences to manage soil and subsoil resources and risks, with a focus on sustainable development, map production, 3D modeling, and the development of digital tools. ISTO, a multidisciplinary laboratory run by the University of Orléans, the National Center for Scientific Research (CNRS), and the BRGM is historically recognized for its research in metallogeny, porous media, magmatology and volcanology, environmental sciences. Recently the laboratory has been also focusing on geodata and digital geology applications.

During this thesis, I was jointly hosted by the two institutions. At BRGM, I was part of the TIA (Traitement par Intelligence Artificielle) working group, which focuses on developing and applying advanced artificial intelligence and scientific computing technologies to process geoscientific data. At ISTO, I was affiliated with the Metallogenic Systems group, which specializes in studying and modeling geological environments and the key processes governing metal concentrations at various scales. The research project received partial funding from a ministerial grant, the BRGM, and the ANR project No. ANR-22-CE56-0001-01, supported by the French National Research Agency (ANR). Additionally, this research initiated an international collaboration with the 3D modeling group of the Canadian Geological Survey (CGS) culminating in a one-month visit to their headquarters in Ottawa, Canada. This mission was partially funded by an Erasmus+ grant. This research work also builds on existing collaboration between the university of Orleans, and the loop consortium* "Three-dimensional Bayesian Modelling of Geological and Geophysical data" (ARC LP210301239).

Manuscript structure

The remainder of this manuscript is structured as follows:

Chapter 1: offers an overview of the current state of 3D geomodelling science, discussing how knowledge is employed and the potential applications of formal knowledge within this field.

Chapter 2: delves into the first area of research, which focuses on formalizing the interpretation process. The methodology and results obtained are presented and analyzed in detail.

Chapter 3: explores the second area of research, which centers on developing an ontological framework to structure expert knowledge. The key findings are presented and further discussed.

General discussion: a section to discuss the main findings of this work, key challenges and limitations, and its position in the general framework of the 3D geomodelling field.

Conclusion: provides a summary of the main contributions of this research, along with potential future directions and prospects.

* https://loop3d.org/

Chapter 1: State of the science

Contents

This chapter aims to provide an exploration of the current state of the science in the 3D structural geomodelling field, with a particular focus on how knowledge is integrated (section 1.2). It has also for aim to discuss the current approaches for modeling expert's knowledge used in the geomodelling process via formal languages –where such approaches are present– (section 1.4). To facilitate a comprehensive understanding, I will introduce key foundational concepts prior to each main section (sections 1.1, and 1.3), providing the reader with essential backgrounds to fully understand the discussions. I will conclude this chapter by a methodological discussion (section 1.5), critically assessing the core challenges addressed by the proposed approach and outlining how these challenges will be addressed.

1.1.General background: geology

Since this research focuses on integrating geological knowledge and expert's experience into the structural modeling framework, an understanding of geology –specifically, of the natural phenomena and how geologists perceive and interpret these phenomena– is essential. This geological background, combined with existing advancements in 3D modeling and knowledge formalization, will help readers recognize the gaps that our approach addresses and understand the primary motivations behind it. To do that, I will first introduce some key concepts related to structural geology and the main geological entities (e.g., objects and structures) that are the focus of this PhD research and are commonly addressed in 3D structural modeling applications. Following this, I will explore how experts naturally build their understanding of these geological entities. Finally, to illustrate how these concepts are applied in practice, I will provide a brief overview of a field experience conducted amid this PhD research.

1.1.1 Structural geology and geological entities

Structural geology plays a crucial role in the study of Earth's history as it a gives a framework for analyzing deformation and paleostress, making stratigraphic interpretations, studying tectonics, and understanding Earth's dynamics (Burke, 2011; Ramsay and Huber, 2002). Structural geology has direct implications on economic and societal challenges, such as in mineral and energy resource exploration, aquifer management, hazard mitigation, and waste management. It helps identify potential zones of mineralization, hydrocarbon accumulation, water resource patterns, active geological zones, and safe storage sites (Bense et al., 2013; Branquet et al., 1999a, 1999b; Fossen, 2016; Levens et al., 1994; Raza et al., 2019; Segoni et al., 2020; Stead and Wolter, 2015, 2015; Tsang et al., 2015). Structural geology takes interests principally in studying the architectures of geological entities –principally geological units and geological structures– and their configurations including their geometric and structural characteristics.

1.1.1.a Geological units

The fundamental building blocks in structural modeling are geological units. Their arrangement, organization, and structure define the geological architecture of a given area. Geological units are distinct and identifiable rock bodies or assemblage of rocks that share common properties and are distinguishable from surrounding rocks. Geological units could be classified based on their various properties such as lithology, age, geometric characteristics, and spatial arrangement (Ashley et al., 1933; Hollis D. Hedberg (2), 1958; Le Maitre, 2004). The geometric characteristics define the shape of the unit, its extension, and localization in space according to some reference (e.g., the north for strike and the horizontal plane for the dipping) (Coe, 2010; Pollard and Fletcher, 2005). Based on these geometric properties, geological units can be classified into two categories: stratified and massive units as sketched in Figure 1-1.

- Stratified units: characterized by bedding and result from stacking geological materials in the depositional environment. Due to gravitational forces, stratified units are deposited in roughly horizontal position as layers. Stratified units generally have a plano-convex shape, which is especially evident in smaller depositional environments like lakes. This shape can also be observed in large-scale marine deposits, which are often described as well-layered due to their relatively uniform thickness over a broad horizontal area.
- Massive units: characterized by the absence of bedding and result from the unsorted accumulation of geological materials. Massive units are associated with tectonic activity, in a result, they can form in various, often non-preferential directions and are typically associated with complex and irregular geometries.

The geometries of units are inherently linked to the processes and conditions implicated in their genesis. While stratified units are associated with sedimentary environments (Boggs, 2006), some igneous bodies are considered as part of stratified units, or more precisely in a stratified sequence, when they are in a paraconformity situation without disturbing the general layering of units, or when deposited in a sequence themselves (Myers, 1975; Tait and Jaupart, 1996). In addition, metamorphic units that exhibit beddings are considered as stratified units (Dmitrieva and Nozhkin, 2012; Fry, 1997; Olszewski Jr., 1980). Massive units generally have igneous origin (Gill and Fitton, 2022; Le Maitre, 2004), but some sedimentary bodies are also considered as massive such as salt domes regardless of their internal stratifications (Fredrich et al., 2003; Jackson et al., 1994). If they don't experience differential deformation, stratified units preserve the homogeneity of their spatial properties (e.g., dip) over their entire volume. Massive units are generally characterized by considerable heterogeneity of their spatial properties, except of some simple shape bodies like sills and dykes. These properties have direct implications on the understanding and interpretation of the geometries of these units during the modeling process.

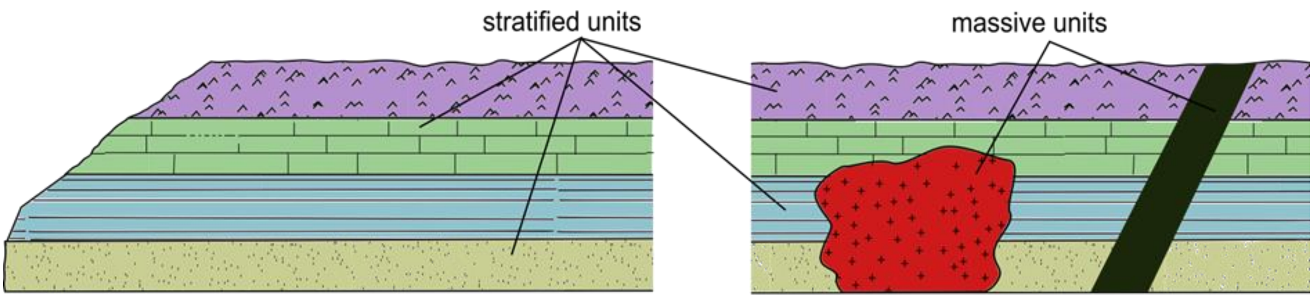


Figure 1-1 Schematic diagrams depicting the two main types of geological units: a series of four stratified sedimentary units (on the left of the figure), that are intruded by two massive units (on the right of the figure)

1.1.1.b Geological structures

Geological structures refer to the patterns occurring on a giving rock body (foliation, fold, dilation bands), or between rock bodies (contact, fracture) (Hills, 1970; Nabavi and Fossen, 2021; Ramsay and Huber, 2002; Torabi and Berg, 2011). Geological structures are dependent on the rock bodies that host them. The current PhD and presented demonstrations concentrate on metric to kilometric structures with particular interest to contacts, folding and faulting structures.

Contacts

A contact between geological units is the shared interface that separates them. Contacts can be classified as either primary or secondary, depending on when they form relative to the unit formation. It is important not to confuse contacts with boundaries, which are the external limits of a unit in relation to its surrounding environment.

- Primary Contacts: Primary contacts form during the initial genesis of geological units. In stratified units, two types of contacts are typically recognized: the bottom and the top. The bottom contact is the interface between a unit and an older unit beneath it, while the top contact represents the interface with a younger unit that overlies it. In contrast, massive units, such as those formed by intrusive processes, have a single contact with the older rock body they intrude into. For an intrusive body the contact is primary, and secondary for the host rock. If massive units are deposited, they can be characterized by top and bottom contacts, similarly to stratified units.
- Secondary contacts result from the modification of the geometry of existing geological units. Three types of secondary contacts exist: erosion contacts, intrusion contacts, and fracturing contacts. When a geological unit is eroded it has a new boundary, that becomes a contact with subsequent deposition of younger units. If erosion occurs from the top, the original top boundary is replaced by an erosional surface. An erosion event can sometimes divide the eroded unit into distinct parts, each with new boundaries. When a unit is tilted or folded, the exposed sides may undergo erosion, potentially leading to partial or complete reformation of boundaries as the unit is fragmented. This specific deformation pattern results in a contact of the younger unit with possibly multiple older units and not only one. Intrusive contacts occur when a new geological unit intrudes into an existing one, altering the original contacts of the host unit by adding new contacts and possibly newer complete boundaries. Finally, fracturing contacts represent the resulting surface when a geological unit is fractured. If the units are not displaced, it is often not considered a new contact or boundary. However, if displacement occurs, it forms a fault, bringing different parts of geological units into contact, as explained further.

Folds

Fold are ductile deformation structures that could occur on geological units in result of the application of differential forces. They are characterized by a bending or curving of geological units (also surfaces) (Figure 1-2). Folds are characterized using various geometric and structural properties, describing and limiting zones of the geological unit on which the fold is occurring:

- Hinge Line: The line that connects points of maximum curvature along the fold. It may be straight or curved, depending on the geometry of the fold.
- Axial surface: is the imaginary surface that connects all the hinge lines of folded layers within a fold. It can be planar or non-planar, depending on the nature of the fold. For example, in cylindrical folds, the axial surface may be planar (referred to as the axial plane), while in more complex folds, it can be curved.
- Limbs: a fold has exactly two limbs representing the two sides of a fold that are separated by the axial plane. The limbs are geometrically described throughout their strike and dip.
- Fold axis: it is defined by the intersection of the two limb planes.
- Amplitude: the maximum perpendicular distance from the hinge line to the plane defined by the inflection points of the limbs.
- Wavelength: represents the horizontal distance between two successive points of the same phase on a wave, such as two consecutive crests or inflection points.

Based on these properties and parameters folds are classified into different types (cylindrical/non-cylindrical, synform/antiform) (Nabavi and Fossen, 2021; Ramsay and Huber, 2002). When numerous folds are arranged in a linear or curvilinear alignment in a defined area, they are generally referred to as a fold train. These folds often share a common orientation and display close geometric features.

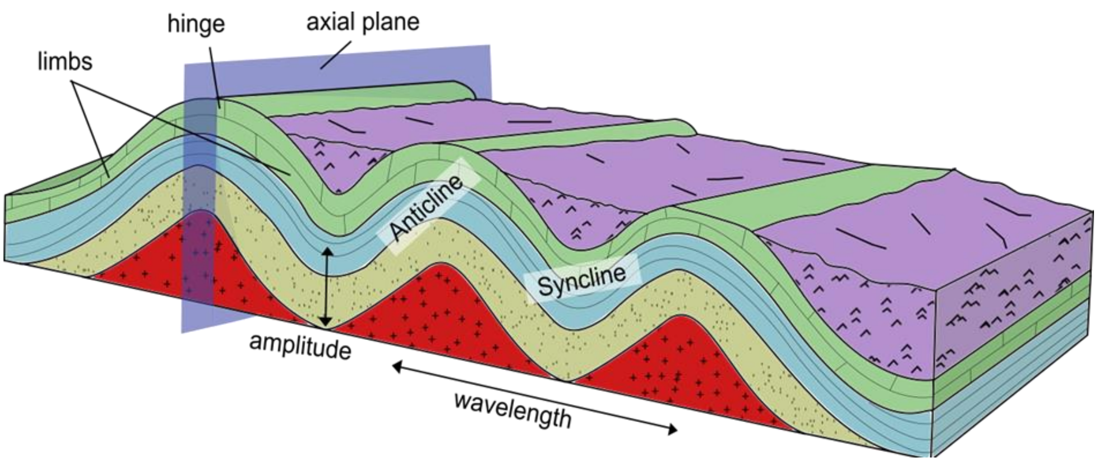


Figure 1-2 Schematic diagram showing a series of folds and their structural parameters

Faults

Faults are brittle tectonic deformation that occur on rock bodies resulting in fractures characterized by relative displacement of rocks on both sides of a fault. If tectonic forces are concentrated along a single surface, it often results in a neat fault with a clearly defined fault plane. However, if the forces are distributed across a broader zone, it leads to the formation of a fault zone as sketched in Figure 1-3. A fault zone consists of several components, each playing a distinct role in the mechanics and geometry of the fault:

- Slip Surface: refers to the specific plane along which sliding, or displacement has occurred during faulting. It may coincide with the major fault plane in some cases, but in others, there may be multiple slip surfaces within the fault zone.
- Fault core: the core is the central volume of the fault zone where the most significant displacement has occurred. It typically consists of a highly chaotic mixture of rock fragments derived from the adjacent rock units. The fault core can vary in width from a few millimeters to meters depending on the faulting conditions and the nature of units. The central slip surface of the fault core represents the major fault.
- Damage Zone: The damage zone is the volume that surrounds the fault core and consists of fractured and faulted rock that has experienced strain but has not undergone the same level of displacement as the fault core. Fractures and faults in the damage zone are smaller and considered minor compared to the major fault.

Faults can be classified based on various characteristics and parameters, including the orientation of the fault plane, the direction of movement along the fault, and the nature of the displacement (e.g., normal faults/reverse faults/ strike-slip faults; thrusting folds/detachment faults).

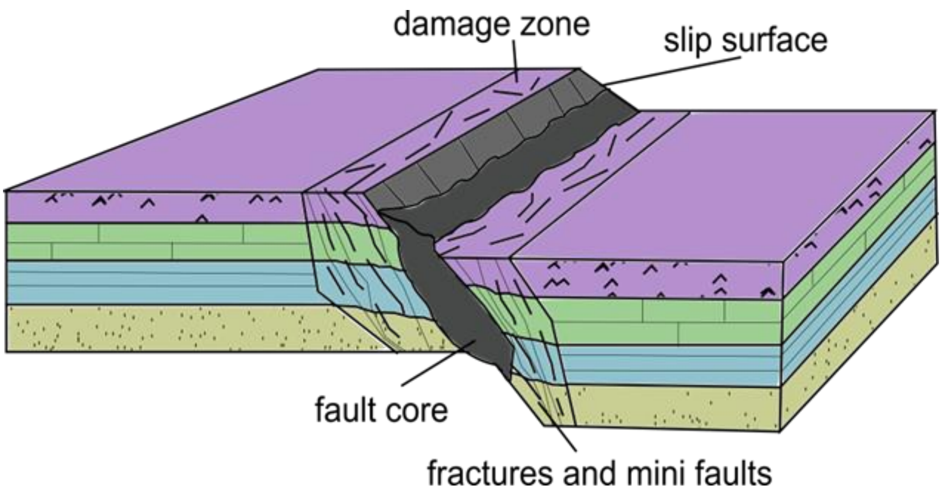


Figure 1-3 Schematic diagram depicting a fault zone with its main components

1.1.2 Thinking like a "naturalist" geologist

Studying the architecture of some part of the Earth calls for developing an understanding of the geometric, topologic, structural and in general geologic properties of the geological phenomena in question. This understanding is inherently related to the conceptualization that is about the mental creation of theoretical constructs (concepts) (Middendorp, 1991; Sequeira, 2014) about geological entities and defining their relations within a mental model (Shipley et al., 2013). To achieve this, experts must gather and process a range of information, interpreting it as necessary, all drawing upon their knowledge.

1.1.2.a Interacting with geological phenomena

Information can be acquired via the observation and measurements of geological phenomena through various types of surveys, such as field, geophysical, and airborne surveys, or by reviewing scientific literature and analyzing pre-existing data in databases. In this section, I will focus on fieldwork, as it is a fundamental and historically significant method for studying geological phenomena and collecting information, and I quote *[…The reading of books and the study of specimens will never make the geologist; the geologist is made in the field, not in the laboratory…]* (Himus and Sweeting, 1952). This does not imply that geologists do not naturally interact or that they employ less knowledge during other types of surveys, but the fieldwork exemplifies how experts interact in a tangible manner with geological phenomena, to develop their understanding and build their mental models.

In the field, experts identify key geological entities through patterns and map them. Mapping often involves measuring various properties (e.g., coordinates, dips, strikes, thicknesses, lengths) using instruments such as compasses, rulers, or digital devices. These measurements are considered conventionally as the abstract raw data (Brodaric et al., 2004; Calcagno et al., 2008), which serves as crucial ingredients for further interpretation. Field experts usually adhere to standard methods of measurement and data collection, acknowledging the absence of epistemic uncertainty in the data, but recognizing aleatory uncertainty due to measurement conditions or instrument precision (Beven et al., 2018; Bond, 2015). In other words, experts trust that a dip measurement is a true dip measurement and not confusedly recorded as a thickness –excluding potential clerical errors, not subjective choices. Detailed notes are taken to document these observations and measurements, along with sketches and photographs to capture the context and highlight significant features that may be challenging to describe in words or through structured data. As it is impossible to include everything observed, experts also rely on their memory to retain significant details of the geological phenomena.

Once acquired, these raw data are interpreted to conceptualize subsurface structures, develop digital models, and for numerous other geoscientific applications. Interpretation aims to transform these abstract measurements into meaningful information about the geological phenomena. Therefore, this process requires a deep understanding of the geological context, as well as the ability to recognize patterns, trends, and anomalies within the data and the use of knowledge. Experts must infer the relations between various geological features, hypothesize about their formation processes, and make predictions about subsurface configurations.

According to Frodeman (2014), observing the natural phenomena itself cannot be dissociated from the interpretation as experts often encounter obscured outcrops, such as rock layers covered by vegetation, weathered soils, erosion, or urban development. In these situations, geologists must interpret the hidden parts to mentally reconstruct the complete picture, to be able to take further measurements or to draw a sketch. Even in cases where geological entities (e.g., units) are clearly visible, experts must still propose interpretive hypotheses regarding the spatial and temporal relations between these entities, as such properties are not directly observable (Brodaric, 2012). For example, observing one geological unit above another is an observation, but recognizing that the upper unit is younger and identifying it as a specific formation is an interpretation. Measuring the properties of this unit results in observational data that is recorded as raw, but such data is itself based on prior interpretation. Therefore, even raw data in geology are inherently shaped by interpretive judgments, as they reflect the observer's assumptions about the geological context.

Whether observing or interpreting, experts rely on their knowledge of geological entities, their characteristics, and their interactions based on established geological principles and rules, such as the principle of superposition (Steno Law) (Catuneanu et al., 2011; Pollard and Fletcher, 2005). The main motor to exploit this knowledge is geocognition.

1.1.2.b Geocognition

Geocognition –as a compound term– is recently emerging in the geoscientific literature (Clary et al., 2009; Libarkin, 2014, 2009; Maeng and Lee, 2015; Petcovic et al., 2009). It refers specifically to the cognitive abilities and activities that geoscientists use. Etymologically, it combines "geo," referring to the Earth, with "cognition," that relates to the mental processes involved in acquiring knowledge and understanding. Cognition is inherently linked to reasoning, which involves the mental manipulation of information to draw conclusions, make inferences, evaluate arguments, and solve problems (Manktelow, 1999; Walton, 1990).

Cognition is a subject of study across various disciplines, including psychology, neurosciences, linguistics, and philosophy. Consequently, different scholars approach the study of cognition and related mental processes from diverse perspectives. For example, Newell (Newell et al., 1972) considers cognition as the process responsible of receiving, transforming information and store it in the mind. Pylyshyn (1989) argues that mental processes, including cognition, thinking, and reasoning, operate on internal representations of the world, indicating that cognition is fundamentally about the mind representing and processing information. For Möystad (2017) cognition is the usage of all conscious and subconscious mental activities to acquire, store, retrieve and process information. In essence, mental processes involved in acquiring, processing, and storing information, that are used to understand and interact with the exterior environment.

Geocognition, a specialized form of cognition, extends these cognitive abilities to the spatial dimension, placing greater emphasis on interpretation to deal with complex geological phenomena.

1.1.2.c Geocognition: a unique type of cognition

As in any scientific discipline, experts in geology work, think and use their cognition in a framework defined by the tasks they want to do. According to leading authors in psychology like Kuhn (2010) and Zimmerman (2000) the cognition in scientific studies, has been intensively shaped by the scientific method, especially through pedagogic education. It has resulted in what is known as scientific thinking and scientific reasoning (Díaz et al., 2023). This implies that scientists are more likely to use their cognitive abilities to systematically observe, hypothesize, experiment, and analyze outcomes in a rigorous and structured manner. They are generally compelled to pursue verifiable and substantiated answers, reflecting the inherent commitment to empirical evidence. The framework of the scientific method influences what problems scientists see, how they solve them, and how they interpret the results (Kuhn, 1997). Thus, scientists may be predisposed to certain methodologies, theories, and interpretations, constraining their cognitive flexibility and potentially limiting the range of inquiry.

> *[…An outcrop is a system of signs the meaning of which is not apparent, but must be deciphered…]* (Frodeman, 1995)

> *[…after all, the 'right answer' is not straightforwardly 'out there' awaiting its revelation…].* (Bond et al., 2011)

For Frodeman, while geology does share the use of the scientific method with other scientific disciplines (e.g., in laboratory analysis), it has a distinctive model as it shares the interpretative and narrative side with humanities (e.g., describing the history of an outcrop). His pioneering works (1995, 2014) qualify geology as a hermeneutic and historical discipline with an important fallibilistic component where interpreted geological information is subject to revision based on new evidence or other interpretations. The work of Bond et al. (2011) aligns with Frodeman's finding as it emphasizes on the acknowledgement of the non-uniqueness of interpretations of the same geological set of information, depending on the interpreter and the conditions of interpretation. Thus, a professional geologist –when it is empirically impossible to prove– must seek the most suitable explanation rather than the right one.

> *[…The complexity of the problems to solve often obliges geoscientists to work as rigorous scientists and as creative artists at the same time…]* (Dell'Aversana, 2013).

The nature of geological phenomena presents a unique framework that fosters the development of distinctive cognitive abilities among domain experts. Unlike in other disciplines, geology experts are often required to study complex systems over various spatial scales with rarely exhaustive informational frameworks. Additionally, geological phenomena are inherently dynamic, shaped by various processes such as tectonics and erosion. This complexity calls for a multidisciplinary approach, drawing on fields such as geophysics, geochemistry, geobiology, and geomathematics to form cohesive interpretations. Consequently, geoscientists develop unique cognitive abilities that enable them to integrate multi-scale, multi-parameter data effectively. These abilities, however, are not only technical but also deeply personal, shaped by each geologist's own approach to problem-solving and interpretation.

1.1.2.d Main geocognitive activities

Despite the quite recent use of the expression geocognition in the geoscientific literature, the geoscientific, psychologic, and cognitive sciences' communities have made a long journey in spotlighting and bringing out the main cognitive abilities and the way they are used by experts, especially in the field. According to various authors (Kastens and Ishikawa, 2006; Manduca and Mogk, 2006; Mogk and Goodwin, 2012; Petcovic et al., 2009; Shipley et al., 2013), the main cognitive abilities of structural geologists are:

- Observational skills: these aspects involve the ability to observe and record details using perception about the geological phenomena noting patterns in rock types, details, variations, structures, geometries, and any other relevant geological feature. Effective observation forms the foundation for all other cognitive processes.
- Spatial reasoning: this aspect enables experts to mentally manipulate and visualize geological phenomena in three-dimensional space, facilitating the interpretation and understanding of structures, geometries, and structural relations.
- Causal reasoning: this refers to the ability to identify cause-and-effect relationships between geological processes and resulting objects and features.
- Temporal reasoning: it refers to the ability of geologists to understand and interpret the temporal aspects and establish chronological order of geological events.
- Analogic reasoning: refers to the ability of drawing comparisons between similar geological phenomena or processes to facilitate understanding and infer relationships.
- Problem-solving: it refers to the mental process of identifying, analyzing, and solving problems, for examples anomalies in the data, or anomalies in the observations (e.g., finding an expected structure within a given context). Besides using the reasoning aspects described above, experts use logical reasoning to solve problems. Logical reasoning involves using principles of logic to draw conclusions from evidence or premises.
- Partial and holistic thinking: experts use both partial and holistic thinking approaches to understand geological systems. Partial thinking is about focusing on specific details or components of a geological system to break down complex problems into manageable parts. Holistic thinking involves considering the system as a whole and understanding how its parts interact influence each other.
- Prediction: this aspect is about using available information to anticipate future events or outcomes based on patterns, past experiences, or existing knowledge. It draws on various cognitive processes such as perception, memory, reasoning, and decision-making. In a sense, prediction is used by experts to propose some hypotheses, that will be further confronted to observations or new information.

Employing cognitive abilities is overlapping with two important aspects that influence what expert see, where to look, and what they consider existing. The first aspect is context. In much of the scientific literature discussing the interpretation and understanding of geological phenomena, authors have emphasized the significance of context, whether explicitly or implicitly (Bashore et al., 1994; Bond et al., 2011; Burke et al., 1990; Doveton, 1994; Frodeman, 1995; He et al., 2024; Lamotte et al., 2020; Zhang, 2008). The context involves the global framework of the geological phenomenon, including the geological history, environmental conditions, and tectonic activity. The contextualization helps experts to develop more accurate understanding. Cognitive processes such as pattern recognition and inference are employed to discern meaningful patterns within this contextual framework. For example, when the known context is a highly folded geodynamic setting, the spatial reasoning and prediction of experts will be guided to look for specific patterns within certain ranges. On the other hand, when considering a polyphasic faulting deformation experts will be taken to look for a certain number of patterns corresponding to faulting families.

The second aspect is the considered reality. This refers to focusing on essential aspects of geological phenomena relevant to a specific investigation, rather than attempting to represent all properties (Wellmann and Caumon, 2018). In geological literature, the idea of considered reality often appears implicitly in studies related to interpretation and modeling. Examples of this can be found in phrases like *[…only structures resulting from this process are studied…]* (Michon and Merle, 2003), *[…folds that have a peculiar geometry are not considered…]* (Bastida et al., 2014) , and *[…We will focus here on metric to multi-kilometer scale structures…]* (Laurent, 2013). The notion of considered reality is closely linked to cognitive processes, such as spatial reasoning. For instance, visualizing a deformed layer with a single, simple fault surface is typically easier than imagining a complex unit made up of numerous blocks displaced by a plethora of fault segments. But they are often defined according to goals of studies.

1.1.3 Geological phenomena in the field: a showcase in Brittany

A field campaign was conducted in the first year of this PhD project. The purpose was to provide practical experience aiming to establish a more concrete foundation for the proposed formalism by highlighting these three aspects: geocognition, interpretation, and knowledge application. This campaign was not intended to cover all practices but aimed to offer straightforward discussions at the outcrop on key aspects: (1) how geologists conceive and conceptualize geometries and structural relations, (2) how they adapt to existing information, and (3) how they update their interpretations.

The Central Armorican Domain was selected as an ideal terrain to illustrate these aspects. This area is characterized by structural complexity due to interactions between the basement and a folded sedimentary cover, combined with restricted access to information, as detailed next. These conditions necessitate frequent reliance on interpretation, reasoning, and geocognitive processes, making it an excellent setting for examining the theoretical discussions introduced. Additionally, the applications presented in the following chapters (Chapters 2 and 3) are based on data collected from this terrain.

1.1.3.a Before stepping in: contextualization

Field practices for data collection and investigations generally involve dealing with these aspects (1) how geologists conceptualize geometries and structural relations, (2) how they adapt to existing information, and (3) how they update their interpretations following common principles. However, these practices can vary significantly depending on the expertise, training, context, and resources of the individual geologists involved. In fact, despite the richness of the literature by handbooks and guidelines of field techniques from the perspectives of structural geology and cognitive sciences (Coe, 2010; Frodeman, 2014; Petcovic et al., 2009; Ragan, 2009; Ramsay and Huber, 2002; Shipley et al., 2013) there is conventional or standardized way that geologists must firmly follow in the field and that applied for all terranes. However, what is often conventional before undertaking fieldwork is to conduct a thorough review of the current scientific understanding related to the study area, which helps establish the context and identify data gaps.

For our research, we focus on the Central Armoricain Domain (CAD) in Brittany, northwestern France, within the Armorican Massif. Previous studies have explored the geology of the CAD and its evolution (Bonnefoy et al., 1996; Faure, 2021; Gumiaux, 2003; Pochon et al., 2016, 2018) that are summarized briefly here (Figure 1-4):

- The Brioverian group represents the oldest rocks in the massif, undergoing deformation and low-grade metamorphism during the Cadomian orogeny.
- The Brioverian group is subsequently overlain by lower Paleozoic formations ranging from the Cambrian to the Carboniferous periods.
- This Paleozoic cover primarily consists of mainly terrigenous sedimentary formations including red conglomerates, sandstones, pelites, black shales and limestones in the uppermost sequence. During the Hercynian orogeny, this sequence was subjected to deformation, low-grade metamorphism, and the intrusion of felsic magmas. For our PhD research, we focus on five geological units: *Formation de Pont-Réan, Formation des Grès Armoricains, Formation de Traveusot, Formation des Grès de Châtellier, and Formation de Riadan* which are part of the lower to upper Ordovician. These units are concordant with one another, lying above the Brioverian Group with a basal angular unconformity. They represent key geological units targeted by quarries for granulates.
- The CAD region is geologically separated from other parts of the Armorican massif by two prominent southeast-northwest striking shear zones known as SASZ for South Armorican Shear Zone and NASZ for North Armorican Shear Zone (Figure 1-5).
- During the Hercynian orogeny, particularly at Visean, the region experienced significant deformation characterized by upright folds with moderately dipping limbs, ranging between 30° and 50°, and fold axes oriented sub-horizontally at approximately N290°. This deformation is associated with axial planar schistosity, which becomes more pronounced towards the SASZ. The folds are also accompanied by fracturing, including N150-160-trending dextral and sinistral strike-slip and tear faults with varying horizontal displacements (Figure 1-5).

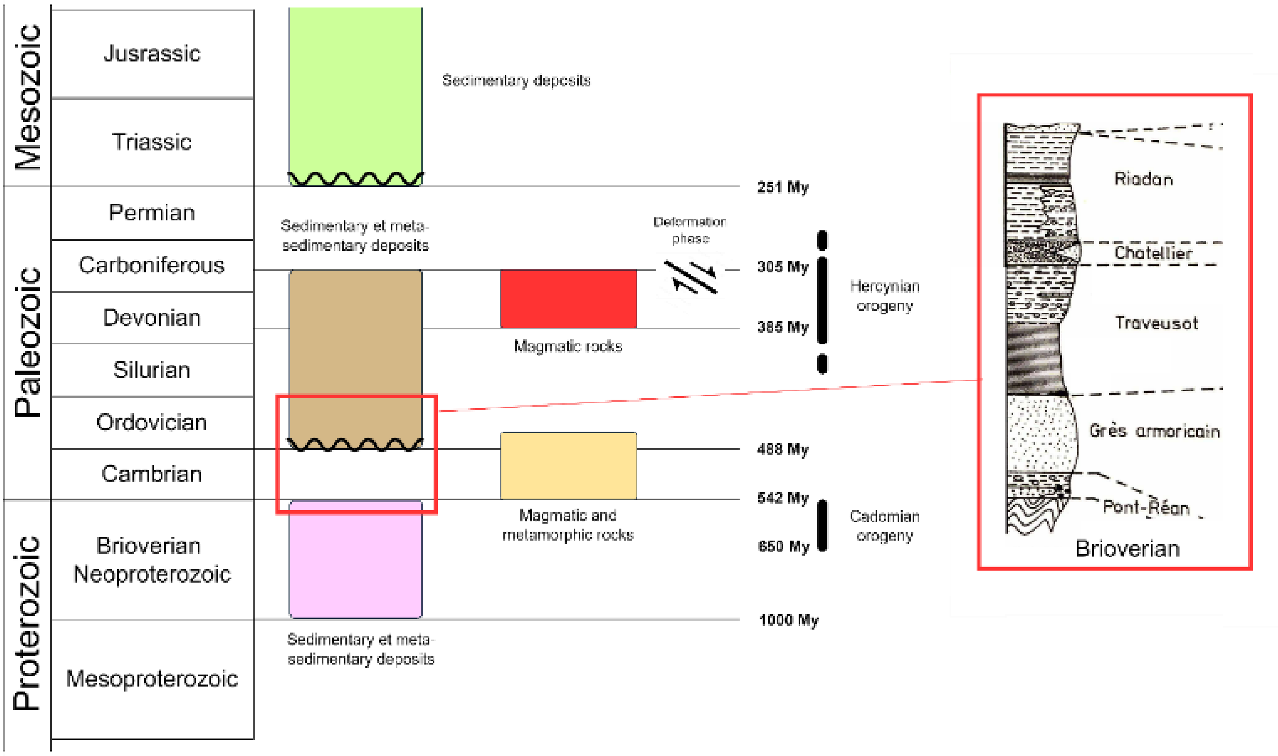


Figure 1-4 Overview of the main geological events in the Central Armoricain Domain, modified after BRGM maps (Bonnefoy., et al 1996)

While the first-order geological framework of the CAD and its evolution are well established, local and second order geometries of units and structures, and their structural relations remain debatable. This uncertainty is justified by the fact that the Brittany region in general, and the CAD particularly, are characterized by an overall flat farming landscape with a dense vegetation cover. Thus, precise data such as field observations are only obtained from accessible locations, mainly through private quarries (Figure 1-5). Ergo, structural interpretations lean heavily on low resolution geophysical data. The limited access to the outcrops has direct implications on considerations taken by geologist on the field, thus the interpretation and the use of knowledge.

1.1.3.b Stepping in: observation and interpretation

Conceiving geometries and structural relations.

In the compact, tight rocks of the *Formation des Grès Armoricains*, faults are primarily observed within fault zones, as shown in Figure 1-6A. These zones are characterized by distinct fracturing, minor displacement, and an increased frequency of faults toward the interior of the faulted area. In contrast, the *Formation de Traveusot,* characterized by altered schist sandstones, also exhibits faults within fault zones, but, these zones are more heavily damaged and often form flower structures with a clearly defined cataclastic core zone, as depicted in Figure 1-6B. The faults in this formation contribute to the occurrence of disharmonic folding on a metric scale.

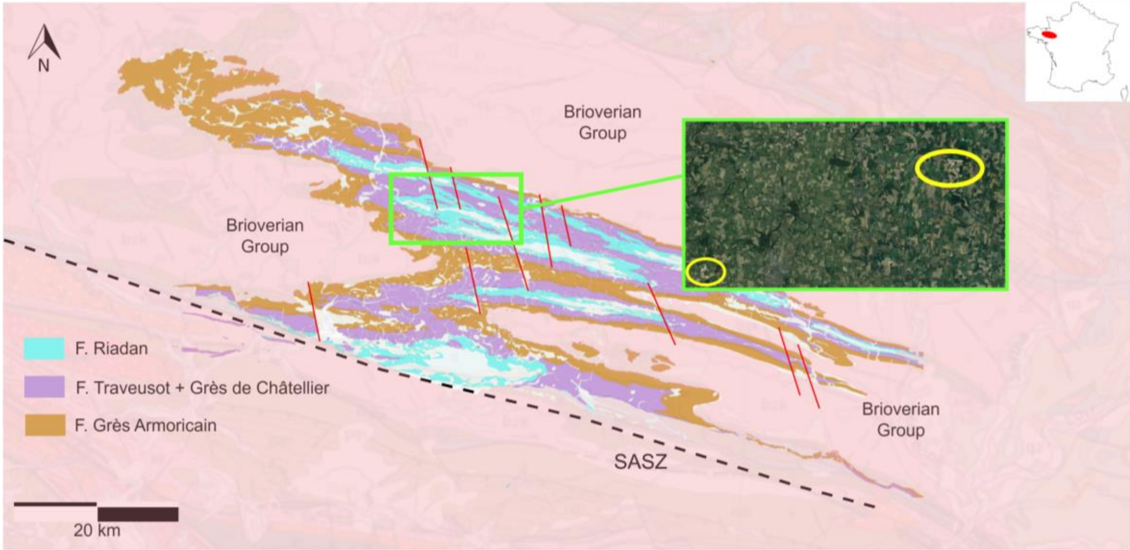


Figure 1-5 Geological map of the Central Armorican Domain. The green frame indicates an aerial image covering approximately 200 km², highlighting the region's extensive farming lands and the limited area (yellow circles) where Ordovician rocks are exposed (rock quarry) origin of the image Geoportail. The three highlighted formations represent main Ordovician units overlaying the Brioverian group. The dashed line represents the South American Sheer Zone (SASZ). The highlighted formations are extracted from the 1:50,000 geological map (BRGM), while the background is taken from the 1:1,000,000 geological map (BRGM). Discrepancies between the boundaries of the units are due to differences in the resolution of the two maps.

Faults in *the Formation des Grès Armoricains* is a convenient framework to apply an analogic reasoning as the geometry of faults is homogeneous (shape and orientation). If someone wants to take some measurements, it could be practical to measure a certain number of faults and extrapolate the results on the remaining ones that are observed partially or holistic but not meticulously measured. It is also possible –depending on the study interest– to ignore some of these faults and focus on a few of them. The consideration and overlooking of some faults are examples of simplification and adaption taken subjectively by the interpreters. On the other hand, when observing faults in the *Formation de Traveusot*, the analogic reasoning is not useful as they are generally heterogeneous. Subsequently, understanding the geometries of these faults require individual attention.

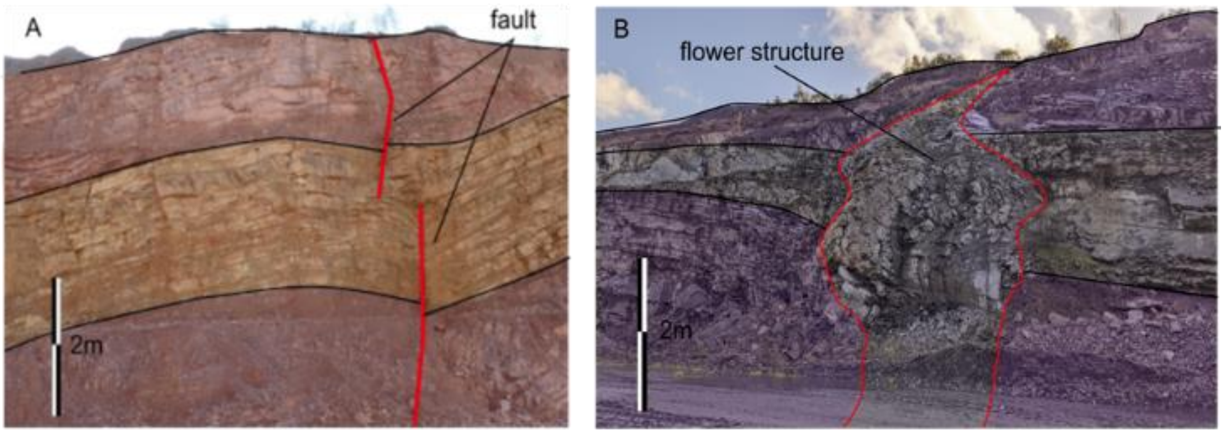


Figure 1-6 Outcrop images showing two types of second-order structures: (A) Formation des Grès Armoricains affected by small and thin faults, (B) Formation de Traveusot with wrench faulting and associated flower structures

When conducting observations to structurally model the studied units in 3D, it is crucial to consider the order of deformation and resulting scale of the structures. Field experts can usually distinguish between primary structures (large-scale folds and original stratigraphy) and secondary structures (like disharmonic folds and foliation associated with folding) (Figure 1-7). This distinction allows for a comprehensive understanding of both types of geometries by applying detailed (partial) and overarching (holistic) reasoning. However, in working stations, unclear observations regarding the structural order can lead to significant inconsistencies in the modeling process and complicate the interpretation of the units' geometries. This issue is particularly problematic when the modeling is performed by experts who were not present in the field. As illustrated in Figure 1-8, observations recorded in survey databases often overlook the importance of scale and order of deformation, making it necessary for experts to address these gaps and refine the datasets based on subjective choices of massaging inputs for 3D modeling.

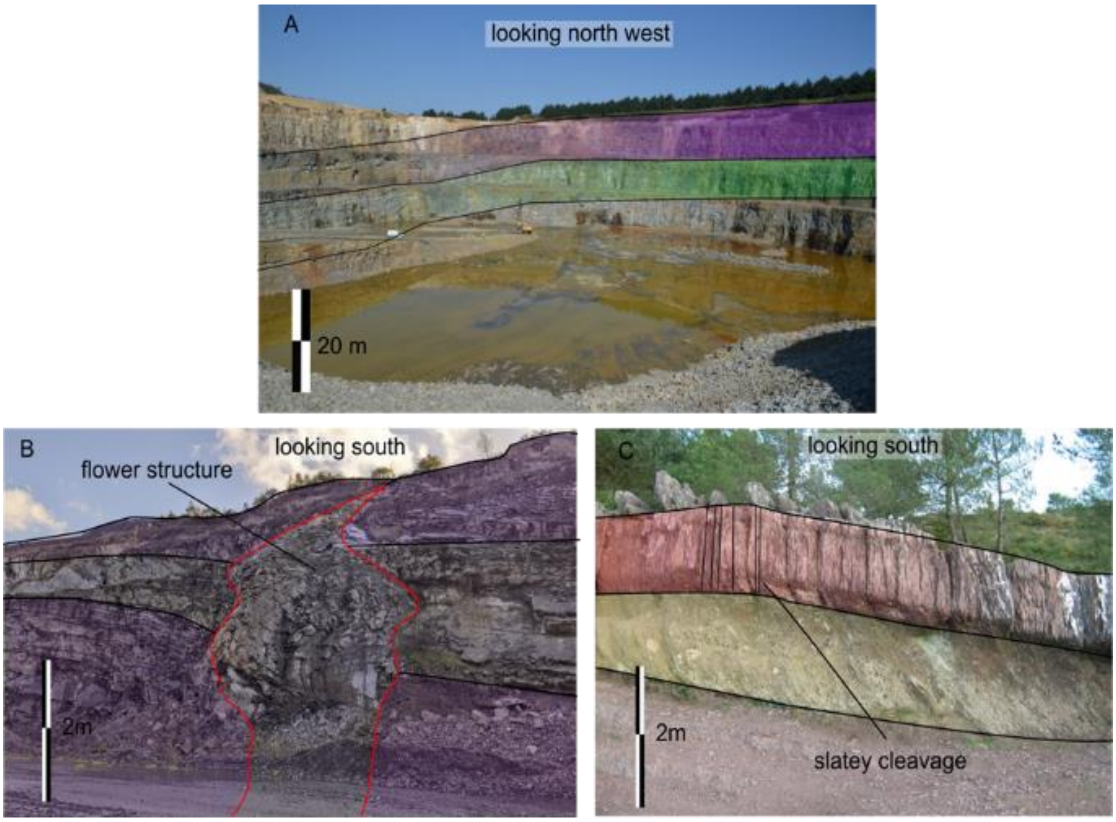


Figure 1-7 Outcrop images comparing first and second order structures: (A) first-order folding structures in the *Formation des Grès de Chatellier*, and (B) and (C) second-order structures. (B) illustrates parasitic folding associated with flower structures in the *Travesot Formation*, while (C) shows slaty cleavage perpendicular to bedding in the red sandstones and pelites of the *Pont-Réan Formation* just above the unconformity with the *Brioverian* basement (not visible here)

Adaptation to existing information

As the targeted geological units for exploration are overlain by younger units, stone quarries in Brittany –the primary source of precise information– are mostly located in the zones of anticlines that offer easier access these units for due to the aging-younging properties of folds. The understanding of the geometries of these anticlines in the subsurface is always updated with the progress in extraction as more outcrops are exposed.

On the other hand, the focus only on anticlines results in a lack of precise information in the syncline zones. This imbalance in observations introduces uncertainty within the interpretative frameworks, such as determining what structures are present and where they are located (epistemic uncertainty). When being in the field and aiming to mentally understand the geometry of this terrane, spatial reasoning is called for. Also, the analogic reasoning is employed to characterize synclines based on the geometries of anticlines. The important aspect that emerges here is the acknowledgement of the uncertain interpretational framework showcasing the inherent fallibilistic component of the geological information.

Updating established interpretation

The Cadomian unconformity represents a good example of how interpreted geological information is subject to debate and update. The unconformity is represented by the superposition of the lower Ordovician sediments on the Brioverian group. Despite that it is observed on the outcrop in very few locations in the CAD, the Cadomian discontinuity is drawn on geological maps over the entire area. The location of the drawn discontinuity is interpreted by experts based on available information back in the time of the map establishment. In our field campaign we have visited the Quarry Cardin La Vigne - Bréal Sous Montfort. According to the map (1/50k) the discontinuity should pass in the middle of the quarry. However, this information was never confirmed by observed or measured information as the unconformity did not outcrop. In this quarry significant extraction operations have taken place in the last years which allows the exposition of new outcrops. During the visit to the mentioned quarry, we identified a potential contact of the lower Ordivician on the Brioverian group. These possible findings come to confirm preexisting uncertain interpretations.

id	x	y	original_z	DESCR	azimuth	dip
p1	358957.575	6764064.98	97	Stratification So avec pendage	171	35
p2	360672.976	6764054.41	84	Stratification So a pendage ver	1	1
p3	363718.193	6763614.29	69	Stratification So avec pendage	216	75
p4	350095.708	6766178.05	55	Stratification So avec pendage	204	70
p5	350011.756	6765725.02	44	Stratification So avec pendage	10	65
p6	344818.318	6765931.32	21	Stratification So avec pendage	115	35
p7	344949.899	6766230.43	30	Stratification So avec pendage	203	65
p8	345458.5	6752111.37	63	Stratification So avec pendage	67	35
p9	344359.113	6771057.61	42	Stratification So avec pendage	210	45
p10	345719.759	6770770.83	53	Stratification So avec pendage	27	45
p11	346207.888	6771129.77	34	Stratification So avec pendage	198	65
p12	345669.795	6770624.69	46	Stratification So avec pendage	180	55
p13	347746.204	6761744.55	78	Stratification So avec pendage	71	80
p14	342452.782	6764977.91	29	Stratification So avec pendage	5	70
p15	342401.118	6765058.59	37	Stratification So a pendage ver	86	1
p16	343367.327	6764558.65	63	Stratification So avec pendage	5	30
p17	342793.498	6765666.19	15	Stratification So avec pendage	36	60
p18	346264.17	6765620.37	36	Stratification So avec pendage	162	45
p19	358964.499	6764912.93	90	Stratification So avec pendage	215	80
p20	323429.25	6747487.89	53	Stratification So avec pendage	356	80
p21	331597.061	6744805.62	29	Stratification So avec pendage	209	60
p22	333844.264	6746473.01	39	Stratification So avec pendage	188	60
p23	334421.087	6746590.46	45	Stratification So avec pendage	358	70
p24	328930.842	6749339	72	Stratification So avec pendage	208	45

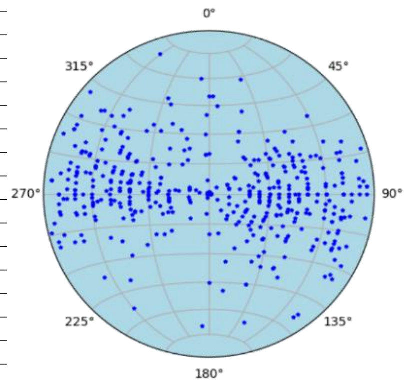


Figure 1-8 Example of field observations from the BRGM geological survey database, illustrating significant variability in structural measurements (a mix of first- and second-order structures). While the measurements target the S0 stratification (original bedding), the large variations in dip and azimuth suggest the influence of second-order folding structures as displayed in the Schmidt Grid

1.2. Structural modeling

3D structural modeling systems are used to build digital representations of geological architectures based on expert's understanding of the subsurface. This PhD research aims to provide original contributions in how these systems operate to assist in interpretation and better utilize expert knowledge during the construction of the digital representations. To clarify how this research achieves that, this section first outlines the main aspects of 3D modeling. It provides an overview of how geological phenomena are represented, various 3D modeling approaches, recent contributions from literature, and discussing how interpretation and knowledge are employed by experts.

1.2.1 Representation methods

The geometric representation, geometric model or digital representation, are terms used to denote the numerical artifact visualized through a computer system, which serves to represent a conceptual model (Perrin & Rainaud, 2013; Wellmann & Caumon, 2018). A conceptual model is a specific instance of a conceptualization (a collection of concepts) structured mentally according to a particular theory (Guizzardi, 2005; Perrin and Rainaud, 2013). In geosciences, the term "conceptual model" is often used not only to refer to a collection of organized concepts but more aligned with the notion of simplification. In Chapter 3, I will present our analysis in comparison with existing literature of these concepts (models and representations). For clarity in the following sections, I will keep using the term conceptual model to describe what experts form in their minds based on their interpretations, that is afterward represented through the geometric representation. Existing 3D modeling tools offer two primary methods for constructing the geometric representation: volume-based methods and interface-based methods.

- Volume-based methods focus on representing geological volumes by discretizing the space into cells such cubic voxel or tetrahedral meshes, and filling them with lithological properties (Maljers et al., 2015; Souche et al., 2013). These methods are highly used in complex sedimentary environments (e.g., channels, estuaries), making them particularly effective in reservoir modeling. Structural surfaces emerge from the resulting geometries of geological volumes and including specific type of structural surfaces such as faults can be challenging, requiring additional techniques, which is why this type of models are not structural models per se.
- Interface based methods focus on representing boundaries of geological entities following the B-rep formalism (a material object is represented by its external boundaries including faces, edges and vertices) (Caumon, 2003; Requicha and Voelcker, 1982). These boundaries that stand for the structural surfaces are either represented explicitly or implicitly. The explicit methods are based on representing interfaces as geometrical objects using for instance depth functions (Hardy, 1971), triangulated surfaces (Mallet, 1992), or parametric surfaces (De Kemp, 1999). On the other hand, the implicit approach is based on using isovalues of a scalar field to represent surfaces (Calcagno et al., 2008; Caumon et al., 2013; Lajaunie et al., 1997). In general, interface-based methods are flexible to be used with various geological settings including sedimentary environments or more complex tectonic settings.

1.2.2 Approaches to structural modeling

Three dimensional structural models digitally represent the current configuration of a part of the Earth that has resulted from a specific geological history, often involving a series of geological events. Because this history is at the origin of the actual configuration of the subsurface, it seems intuitive to digitally model the geological processes and events to build the current configurations, which is known as process-based modeling (Tikoff et al., 2013). In such framework, experts predict the sequential spatial distribution of geological properties and structures using a knowledge framework that models the succession of identified events based on their understanding of the geological history. For example, in a kinematic modeling system like Noddy (Jessell and Valenta, 1996), experts input a set of initial parameters, such as the number of units, deformation structures, their properties, and ages, to forward-compute the outcomes through the execution of identified events. These outcomes are then compared with existing input data (Armit et al., 2012), such as field observations.

This approach is most effective when experts possess a comprehensive understanding of the geology of the area, including detailed account of geometries of geological objects, as well as their spatiotemporal relations throughout their genesis history. However, the complete geological history of an area is often not well known, particularly in tectonically complex settings. The large number of geological and physical factors that influence the geometric evolution of geological elements makes it practically impossible to achieve exhaustive understanding. Additionally, initial conditions, which determine the starting point of modeling, thus, the outcomes, are often unknown. Therefore, relying solely on this approach to produce a final precise geometric representation is challenging due to the multitude of factors that need to be numerically modeled, requiring significant iterative tasks. Yet, this approach remains highly valuable and is better suited to building a global and conceptual understanding. It allows experts to run various forward simulations and explore flexibly various geological scenarios.

Instead of trying to fully understand geological processes and events by experts, modeling them with systems, and then using observations to validate resulting models, one approach is to use the inverse problem framework. In this framework, the goal is to infer the subsurface structures and their properties by working backward from the available observations combined with priors and knowledge (Tarantola et al., 1982). Although significant endeavor is dedicated to developing this framework (Galley et al., 2020; Giraud et al., 2016; Grose et al., 2018, 2019; Iglesias et al., 2016; Tarantola, 2005; Wellmann and Caumon, 2018; Wellmann et al., 2014a), fully solving the problem of reconstructing the subsurface configurations within the inverse framework remains extremely challenging and ill-posed (Wellmann and Caumon, 2018). This is mainly due to the non-uniqueness of solutions, where multiple models can fit the same observed data, the dimension of the problem in terms of the number of required parameters to infer, and the sensitivity to errors, where small inaccuracies in the observed data –often the case– can cause significant variations in the resulting model. Accordingly, such a framework requires systems to run intensive simulations with heavy computation and best fit sampling tasks. Moreover, formulating geological knowledge itself to integrate multiple types of observed data with their full semantic contents –such as seismic, gravity, and field observations– into a unified model, remains challenging.

A pragmatic solution is to focus on constructing current subsurface configurations based on the geometric and topological properties of geological entities through interpolation of existing information. Interpolation is about estimating some values in unknown locations based on known values and their spatial correlations (Franke, 1982). In this way, experts mentally test geological events and hypotheses that could have led to the present configuration in an iterative process while allowing the algorithms to focus on better describing the geometric implications. This approach can be seen as a simplified form of the inverse problem framework, where experts perform mental simulations of geological processes. However, despite the efficiency of expert geocognitive abilities, their capacity to explore a wide range of possible solutions is limited compared to specialized geomodelling systems designed for inverse problems. As a result, experts often focus on a limited set of solutions, which can introduce subjective biases into the modeling process. Despite this, the interpolation-based paradigm is versatile, accommodating both large and small datasets, heterogeneous inputs, and is adaptable to various modeling contexts. It does not require a comprehensive understanding of the full geological history at once, but rather a sufficient understanding of the present state, although geological history and processes are typically considered implicitly or explicitly introduced in the workflow in the form of restoration or simulation (Laurent, 2013). Therefore, this paradigm is widely adopted by most commercial geomodelling tools such as Gocad/SKUA[1] (Jayr et al., 2008; Mallet et al., 1989), Geomodeller[2] (McInerney et al., 2005), Petrel[3], and GDM[4].

Given that this PhD research aligns closely with how common geomodelling systems handle the construction of geological configurations –though with a greater emphasis on structuring the interpretation process and associated knowledge– I will provide more details in the next section about this modeling paradigm, focusing on the different steps used to build the model and how experts interact, main limitations to incorporate inputs and knowledge, and main advances following this modeling paradigm.

1.2.3 Building 3D structural models

As outlined by Wellmann and Caumon (2018), the construction of 3D structural models involves two key steps: constructing a conceptual model of the studied area through the interpretation of inputs. This model is converted into a parameterizable mathematical model that is executed to produce geometric representations based on interpolation functions. To better focus on the use of knowledge, what modeling tools offer and how experts are involved, I will treat naively at this point, the construction of the conceptual model and the geometric representation as being done distinctively in two stages: (1) producing the conceptual model, and (2) constructing the geometric representation (Figure 1-9). Two additional elements will be discussed as well as they are related to knowledge application and experts' involvement:(3) model validation, and (4) workflows for achieving the final geometric representation.

[1] https://www.mirageoscience.com/fr/logiciels-industrie-miniere/gocad-mining-suite/
[2] https://www.intrepid-geophysics.com/products/geomodeller/
[3] https://www.slb.com/products-and-services/delivering-digital-at-scale/software/petrel-subsurface-software/petrel/petrel-geology-and-modeling
[4] https://www.brgm.fr/fr/logiciel/gdm-suite-gerer-modeliser-visualiser-donnees-geoscientifiques-3-dimensions

1.2.3.a Producing the conceptual model

Conventionally, the conceptual model includes the types and number of geological units, their chronological and spatial orders, and the stratigraphic relations between these units (namely, types of contacts). It also includes the description of major structures present in the study area. The conceptual model serves as a simplified abstraction that is subsequently communicated to the modeling system, though experts often develop a more detailed understanding of the geometries and properties of geological entities throughout the different modeling iterations.

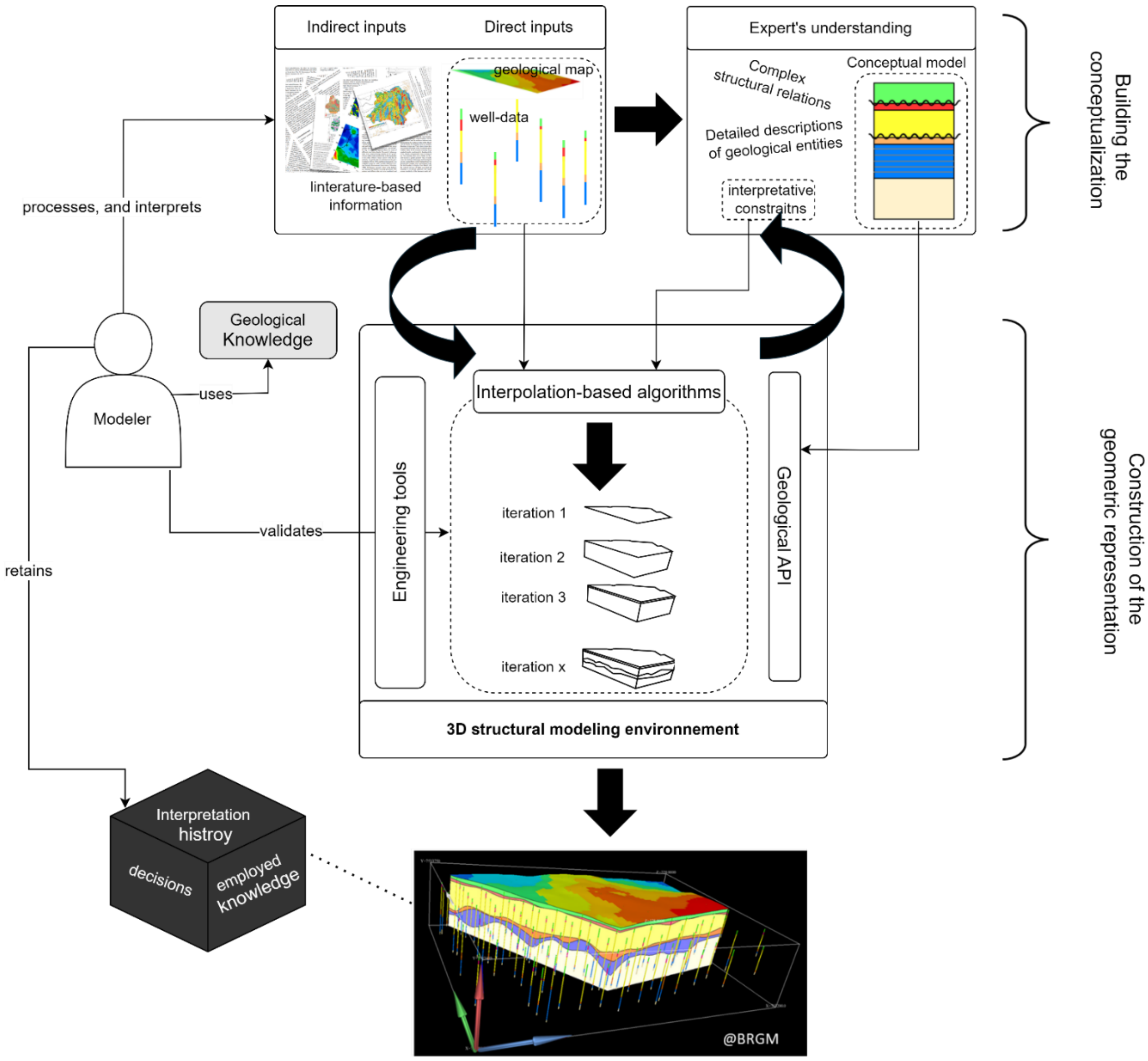


Figure 1-9 Diagram illustrating the common workflow for constructing a 3D structural model using a 3D modeling system based on geometric representation through iterative interpolation. It highlights the points where the expert modeler interacts with the process, managing and interpreting inputs, introducing interpretative constraints, and feeding the conceptual model into the system. The system progressively builds the model, while the decision-making process and interpretation history remain stored in the expert's mind, functioning as a "black box." The final output is a numerical geometric representation of the model

Inputs and contextualization

During this initial stage of modeling, considerations regarding scale and the tolerance for spatial variability of geological entities are established. These parameters are often set to meet the resolution requirements of the project and capabilities of the modeling algorithms. Subsequent interpretation and the generation of geometric representations must adhere to these constraints. While the conceptual model and the experts' understanding evolve and become more refined through iterative processes, the foundational elements of the conceptual model are typically defined from the outset (main geological units, main stratigraphic relations). This is particularly true when inputs include literature-based data and pre-interpreted sources, such as geological maps, interpreted core samples, and cross-sections, which provide an overview of the main geological units, structures, and their spatial and temporal relationships.

Inputs for structural models are diverse, ranging from localized field observations to extensive datasets from large-scale surveys. As depicted in Figure 1-10, depending on their incorporation into structural modeling algorithms, inputs can be classified as either direct or indirect. Direct inputs are those directly utilized by modeling algorithms, such as observations of unit boundaries or geological maps. Indirect inputs, such as well logs, require preliminary processing and interpretation to extract usable information, like delineating unit boundaries in well logs based on the petrophysical properties. Experts massage inputs to integrate them into the modeling system or to mentally interpret them, based on various criteria: primarily the quantity and type of information they can yield, their precision, and their relevance to the specific study context. A significant aspect of this process involves the tacit application of expert knowledge, often without explicit documentation, as these aspects are typically not captured or required by modeling systems (e.g., ignoring field observations with measurements of small-scale secondary structures).

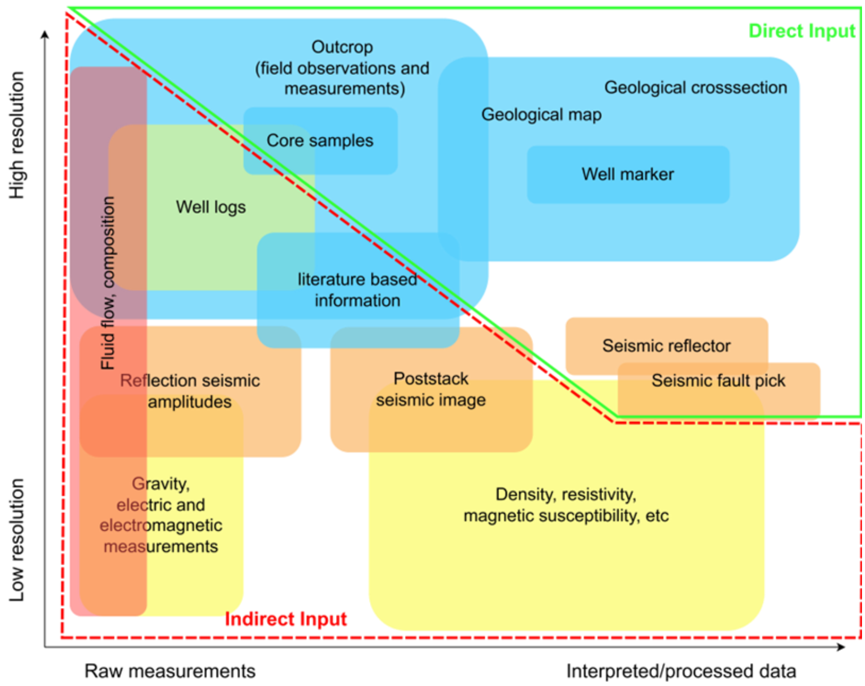


Figure 1-10 Schematic diagram illustrating the different types of inputs for 3D modeling, modified from Wellmann and Caumon (2018). It distinguishes between direct inputs (green frame), which are fed directly into the modeling system, and indirect inputs (red frame), which require additional processing and interpretation before use

Interpretation

Regardless of the type of inputs, experts often engage in mental manipulation of both direct and indirect data, leveraging their knowledge and understanding of geological concepts. This includes their familiarity with various geological phenomena and their ability to recognize how these phenomena are partially manifested in modeling inputs through their properties. For example, sparse observations of the same geological units are commonly encountered due to limited sampling density. When interpreting such situations, experts consider the nature of the unit (e.g., intrusive versus depositional) and apply geological principles (e.g., law of lateral continuity) to hypothesize the geometry and spatial distribution of the units. These interpretations are often made implicitly, without explicitly referencing the underlying knowledge or principles.

The depth of this knowledge depends on the experts' experience and academic background (Bond et al., 2012). Consequently, modeling results are heavily influenced by the experts involved and their decisions, making models inherently subjective (Baddeley et al., 2004; Lacchia et al., 2020; Wood and Curtis, 2004). Experts utilize reasoning, inference, and interpretative processes similar to those employed in fieldwork, as discussed above. This adds another layer of subjectivity to the models, as these geocognitive processes vary between individuals. For example, simplifying a complex fault network by omitting smaller fault segments is a common interpretative practice. One expert might simplify fault geometries in one manner, while another might do so differently, resulting in model variations that are not explicitly documented or justified within the model framework.

The subjectivity inherent in the interpretation process and use of geocognition is not necessarily problematic; in fact, it offers valuable flexibility, allowing experts to think creatively and propose various theories based on limited information. The real issue arises when systems fail to account for these processes, leading to the loss of the interpretative history that experts relied on to develop their understanding. Systems often do not track which information was used to generate a specific interpretation, why a certain interpretation was updated or rejected, or what simplifying assumptions were made. Two interrelated factors contribute to this problem: (1) experts frequently rely on tacit interpretative processes and tasks, and (2) systems are not designed to capture or record these processes. This creates a cycle where the absence of documentation prevents systems from fully supporting or understanding expert reasoning. As a result, systems remain dependent on the final outcomes produced by experts, without effectively contributing to or capturing the underlying rationale behind those outcomes.

1.2.3.b The geometric model: a mathematical framework

The mathematical framework is the core a 3D structural modeling system. It provides three important components: the digital geological module for describing the geological entities that will be modeled, interpolation functions associated with their support (cells, surfaces, scalar field), and the different engineering tools to enable editing and interacting with the geometric representation. Inputs are also described in this mathematical framework through their representation properties (points, surfaces, and volumes).

Introducing the conceptual model

A fundamental component in modeling systems is the geological module that allows integration of different geological units into a stratigraphic column (Figure 1-9), often implemented using relational databases. The stratigraphic column encompasses the main concepts of the conceptual model (properties of units, spatiotemporal relations), which the interpolation outcomes must adhere to. This module also includes major structures such as faults and how they possibly relate to the different structural surfaces in the model, also implemented with relational database. Despite these capabilities to incorporate most aspects of knowledge defined in the conceptual model that have simple overall geometric properties, other concepts remain challenging to integrate.

For example, when faults intersect or abut unit boundaries, or propagate along them in detachment systems, it becomes difficult to capture these interactions accurately. Similarly, configurations involving duplex structures with flat thrusts, ramps and folded layers also present challenges, as multiple fault surfaces often intersect the same unit boundary, complicating their representation in the model. This complexity arises because a surface cannot simultaneously serve as both a unit boundary and a fault object. Moreover, while step-functions in implicit modeling are commonly used to represent faults –effective for faults that cut across stratigraphic units–, faults that run parallel to stratigraphy are particularly problematic to represent (Grose et al., 2021a). Current geomodelling systems thus, lack the necessary infrastructure to capture such geological configurations accurately with the full semantics. Instead, practical workarounds are applied which include modeling the detachment surface as a thin unit representing the fault gouge, often with an artificially assigned age to maintain consistency with the stratigraphic sequence in the former case. In the latter, the solution is to ignore the faults and model only the blocks of the same unit as separate entities with assigned different ages and defining onlap relationships to avoid inconsistencies.

Although these solutions involve modifying existing entities and incorporating new ones, or using conflicting stratigraphic criteria, the real issue lies not in the solutions themselves –since they are necessary– but in the loss of tacit geological knowledge. This knowledge remains with the experts who construct the model, while the systems lack the ability to understand the underlying geological concepts, focusing solely on geometric representation. As a result, subsequent model revisions by different experts can be problematic due to the absence of this implicit knowledge.

In addition to providing framework for the modeled geological entities, the mathematical framework supports a variety of operations based on the relations defined in the conceptual model, including geometric manipulations to apply cross-cutting relations through fundamental techniques such as Boolean operations (union, intersection, difference) and other geometric transformations that adjust surfaces and volumes. Additionally, interactive editing tools enable users to manually or automatically refine and adjust the model. The flexibility in the integration of geological information and in editing the model is based on whether an implicit or explicit modeling approach is used. In implicit modeling, the sequence of geological units is defined through subparallel surfaces, which simplifies the modeling and editing of complex geological scenarios.

Interpolation

To generate a geometric representation, experts select appropriate interpolation functions and execute them, we speak about the realization of the conceptual model. Interpolation calculates the values of geological properties in areas not directly covered by information, using known values from other locations. In this context, interpolation estimates the spatial location of interfaces, thereby predicting the geometry of modeled entities under the inherent assumption of smoothness of the modeled interface. Over the past few decades, numerous interpolation algorithms have been developed by the geoscientific community to enhance the performance and the ability to integrate geological concepts (Franke, 1982; Hillier et al., 2014; Lajaunie et al., 1997; Oliver and Webster, 1990; Setianto and Triandini, 2015; Zhong et al., 2020). For example, kriging (Oliver and Webster, 1990) is a widely used geostatistical method that provides the best linear unbiased estimation based on spatial correlation of data. Radial Basis Functions (Cowan et al., 2002; Franke, 1982) are also popular, offering flexible interpolation through smooth functions that adapt to the spatial distribution of data. The least squares method on the other hand allows for smooth surface generation by minimizing the overall error along the surface curvature (Levy and Mallet, 1998).

Despite their ability to incorporate coordinates and orientation properties, interpolation algorithms struggle to include certain key properties of geological units. For example, using the thickness of a geological unit to understand its geometry might seem intuitive, integrating this concept into the interpolation framework remains challenging (Marinoni, 2003). In implicit representation, Laurent (2016) shows that variations in the scalar field gradient between neighboring points can cause rapid and uncontrolled changes in thickness, which can lead to inconsistences in cases of poor sampling or significant thickness variation scenarios. Consequently, thickness, as a conceptual property that characterizes geological units and can vary or remain constant depending on the unit type, is often fully understood only by experts. Another challenge arises with the use of inner beds observations which are made within a geological unit without directly measuring its boundaries or orientations. Although these observations are very common in field surveys, they remain unincorporated within interpolation algorithms as it is challenging to directly include where the unit shall relatively pass. Some algorithms such as Gibbs sampler (Geman and Geman, 1984) provide a framework to use these interbed information in a simulation framework, though, many commercial tools avoid using them due to the high computational cost and processing time involved.

Specialized mathematical frameworks

While the distinction between the construction of the conceptual model via interpretation and the geometric representation via interpolation is well understood by experts –often intuitively– a potential ambiguity arises concerning the role of interpolation. Interpolation generates information about the geometry of geological units in unknown locations, thereby extending or detailing the conceptual model in specific locations. This raises the question: Is interpolation, initially used to build the geometric representation, also part of the construction of the conceptual model process? If yes to which extent are associated? This subtle but significant questions are further explored in this PhD based on ontological analysis.

From the point of view of the current state of the science, most modeling systems and subsequent advances, focus on improving interpolation algorithms and the mathematical framework to better incorporate geological knowledge, ultimately aiming for more accurate geometric representations. In other words, bringing closer the construction of the conceptual model and construction of geometric representation. For example, various approaches have been developed to encapsulate structural parameters of folds (e.g., fold axis, axial surface, and vergence) to better enhance their geometries during modeling (Laurent, 2013; Laurent et al., 2016; Pizzella et al., 2022), or to take advantage of folding patterns and regularity properties (e.g., wavelengths, amplitudes) to predict geometries of folds using specific functions such as Fourier Series for instance (Grose et al., 2019, 2018, 2017).

Even when these methods are focused on folding characterization, limitations arise when trying to introduce specific instances (e.g., a single fold) at precise locations, because these methods are focused on the global parameterization over the area of the model. One of the most practical approaches to bypass this issue is to include localized interpretations about folds, or any other geological entity, as interpretative constraints such synthetic points or cross-sections (Arienti et al., 2024, p. 200; Caumon, 2003; Lemon and Jones, 2003) in a cumbersome iterative process. This way of introducing concepts as drawing or constraint points, leads to the loss of significant epistemic and fallibilistic properties inherent in expert's interpretations as discussed previously, in addition to the fact that these interpretative constraints will be used merely to force the interpolation, leading to more loss of semantics.

Similarly to folds, faults hold significant importance and pose challenges in structural modeling due to their complex geometries, particularly when multiple segments are involved. Work such as that of Cherpeau et al., (2012) proposes a method to integrate uncertainties related to fault orientation, location, size, and sinuosity in a stochastic framework to better link fault's segments. Godefroy et al., (Godefroy et al., 2019) introduce a approach to integrate prior information such as distances between segments, orientations, structural parameters of segments, and fault families. This method improves the detection of faults within a combinatorial framework by defining more consistent subsets instead of exponentially increasing possibilities when using non-contextualized inputs. These methods act on the identification of faults.

Various other works exist in the literature and are worth noting (Grose and Ailleres, 2021; Huang et al., 2023; Krajnovich et al., 2020; Madani Esfahani and Asghari, 2013; Rotevatn and Jackson, 2014; Wu et al., 2015; Wu and Xu, 2003; Zanchi et al., 2009; Zulauf et al., 2017) though we will not detail all of them here. A common feature among these methods is their focus on improving the mathematical framework and interpolation techniques by incorporating additional geological knowledge, thereby reducing reliance on expert involvement. However, these approaches often concentrate on aspects of knowledge that can be mathematically formalized and spatially described, neglecting those that cannot. As argued in this work, integrating more knowledge in systems to help in the interpretation can significantly enhance their geometric representation.

A critical consideration of this PhD research is the fact that various systems often describe geological knowledge about modeled geological entities differently using specific terminologies, even though this knowledge is largely shared across different 3D modeling applications. Additionally, there is a considerable variability in formats and specific implementations used by different 3D modeling systems to describe resulting 3D models, which often results in incompatibility between systems. For example, one system might employ a unique grid format for storing surfaces, embedding assumptions about cell size, orientation, and interpolation methods. Users unfamiliar with this format may find it challenging to interpret or modify the data outside the original software environment, restricting their ability to interact with or refine the model. This lack of compatibility and standardization impedes data sharing and collaboration among different research groups. Furthermore, the geological knowledge integrated into these systems –such as rules governing fault behavior and layer displacement– is often embedded in the software's backend without explicit reference to the geological principles or physical laws involved. This knowledge is frequently implemented in specific programming languages, making it difficult for users to understand or adjust underlying assumptions unless they possess deep technical knowledge of the software's codebase.

1.2.3.c Model validation and assessment

The validation of structural models is grounded in a combination of expert evaluation and systematic methods to ensure adherence to geological principles, representational integrity, and knowledge about the study area (Caumon, 2010). Expert assessment, carried out mentally during iterative stages or upon model completion, draws upon extensive geological knowledge (Calcagno et al., 2022; Mas et al., 2022; Perrin and Rainaud, 2013; Pimenta et al., 2005; Shragge et al., 2019). Similarly to the choices and decisions made during the construction of the geomodel, many of the validation decisions remain obscure and implicit. Expert's evaluations are complemented by mathematical frameworks within modeling systems, which incorporate checks of some of the topological and geometric properties of the digital models based on their description in the conceptual models, ensuring compliance with principles namely cross-cutting and superposition (Thiele et al., 2016a, 2016b; Zlatanova et al., 2004). Recent work by Parquer et al. (2024) extends this by incorporating additional geological principles, including inclusion and topological rules, to validate models.

For model validation, many geomodelling tools can incorporate additional knowledge not included during the interpolation phase. For instance, properties such as minimal and maximal thicknesses derived from inner bed observations can be used as inequality constraints to refine surface localization (Bourgine et al., 2008; Caumon, 2003). In addition, various approaches are proposed to address the integration of prior knowledge, such as expected orientations of units or fault characteristics, into the validation process. For example, Wellmann et al. (2014b) proposed a fault model validation framework incorporating fault type, layer continuity, and expected structural features like grabens. Although not detailed here, the validation process is further enhanced by probabilistic approaches and uncertainty assessments, particularly those based on inverse theory and Bayesian methods. These approaches refine models by aligning them with observed data and updating uncertainties through iterative adjustments(Chilès et al., 2007; Courrioux et al., 2015; Hasan and Shang, 2022; Heim et al., 2019; Helton and Sallaberry, 2017; Jessell et al., 2018; Kang and Lee, 2020; Khaninezhad et al., 2012; Krajnovich et al., 2020; Liang et al., 2021; Olierook et al., 2021; Pakyuz-Charrier et al., 2018). For a more comprehensive discussion of uncertainties in 3D modeling, refer to Wellmann and Caumon (2018). The key takeaway is that, given the effectiveness of current methods in incorporating knowledge to validate and assess models, it would be beneficial to include this knowledge earlier in the iterative modeling process to aid in decision-making instead of final validation, but in a more explicit way. From another perspective, challenges remain in documenting the validation processes, specific principles or knowledge aspects used to assess, and ensuring interoperability due to different vocabularies and formats, as noted in the section on the mathematical framework. Thus, a more formal way of describing these knowledge aspects can enhance their systemic integration into geomodelling systems.

1.2.3.d Geomodelling workflows

Geomodelling workflows define the sequential steps during the construction of models from inputs interpretation to the generation of the final geometric representation. These workflows differ according to the considered knowledge aspect. For example, it is possible to start by assessing structural properties of geological units over the entire study domain to derive major trends, then refine the model with more local properties (Fernandez et al., 2009; Maxelon et al., 2009). On the other hand, it is possible to prioritize the cross-cutting principle, thus, building objects in the model while considering the chronological order of geological events (Perrin et al., 2005; Perrin and Rainaud, 2013). When using common geomodelling tools, there is a certain consensus: first, modeling discontinuities such as fault networks, followed by the building of layers and their geometries chronologically. This preference arises from the assumption that layers will potentially behave in a more correlated way within the same blocks. While these examples of workflows attest to the consideration of important aspects of geological knowledge, they are rigid and defined specifically to similar case studies. In practice, operators often follow a more arbitrary workflow when interpreting information to build models. This workflow depends on the available data, the quality of the data, and subjective preferences (e.g., top to bottom). However, greater flexibility in workflows can introduce certain challenges. For example, insufficient adherence to geological constraints, (e.g., sequential deposition) may result in inconsistencies that require further correction.

A very important aspect for this PhD research is that most modeling systems do not record the detailed steps and knowledge aspects involved in the geomodelling workflow, even though they provide a framework for it. This lack of documentation makes it difficult to account for epistemic variability, as preferences, decisions, and considerations often go unrecorded. For example, two experts working with the same initial data could produce vastly different models. Documenting the steps and strategies used can help reveal epistemic differences and guide, hopefully, toward a more consistent model. Additionally, if two experts produce similar models but use different methods or decision paths (such as determining which objects crosscut others and in what order), any updates by other experts who are unaware of the original strategies could lead to inconsistencies as they need to re-interpret the information in the geometric model.

1.2.4 A note on Machine Learning methods

An overview of 3D structural modeling methods would be incomplete without acknowledging the recent advancements made by Machine Learning research groups. Machine learning (ML), a subset of Artificial Intelligence (AI), focuses on systems that learn and improve from experience without explicit programming (Carbonell et al., 1983; Mohri et al., 2012; Samuel, 1959). A primary strength of ML is its capability to automate tasks post-training, such as pattern recognition in images. Another advantage is its ability to learn and operate on large datasets. However, this capability also poses a limitation, as substantial data is required to achieve reliable predictions. Ongoing research aims to mitigate this dependency on vast datasets (Kotsiantis et al., 2006; Zhang and Ling, 2018).

In geosciences, ML is employed in areas with extensive datasets, such as pattern recognition in thin-sections and lidar images (Cracknell and Reading, 2014; Dramsch, 2020). In the context of 3D modeling, ML has predominantly been utilized for processing seismic data in the petroleum industry, benefiting from the availability of large datasets (Bai and Tahmasebi, 2020; Di et al., 2020; Dramsch, 2020; Guillen-Rondon et al., 2020; Gunderson et al., 2022). For scenarios involving sparse data, ML applications are still in early development stages, though promising developments are emerging. For instance, Hillier et al. (2023) have explored the use of geological constraints and stratigraphic relations within neural networks, showcasing significant potential in building 3D structural models over large areas using hundreds of thousands of well-picks as inputs.

A critical area of ongoing research is the explainability (explaining the reasons behind the decisions made) and interpretability (understanding the inner workings) of ML-driven applications. This aspect is crucial in structural modeling, as critical decisions may depend on the resulting 3D models. As the one of the main challenges that this research is addressing is the tacit application of knowledge by experts. From this research perspective, although it will help better incorporate more expert knowledge in the process, adopting machine learning for automating interpretive tasks will not however address the explicit aspect of knowledge. Therefore, I believe enhancing the comprehensibility of ML models is essential for their effective integration into 3D structural modeling. Yet, this research does not reject ML applications; rather, it identifies potential synergies between ML techniques and our approach based on knowledge formalization, as discussed briefly in the prospects at the end of this dissertation.

1.3.General background: knowledge engineering

As discussed previously in this chapter, geoscientists –particularly those involved in 3D structural modeling– rely on a complex body of knowledge. Recognizing the need and potential for more knowledge incorporation in digital systems, these experts are increasingly advocating for knowledge engineering and formalization (Abel et al., 2012b; Brodaric and Gahegan, 2010; Mazzetti et al., 2022; Simons et al., 2006; Urvois et al., 2022; Whitehead and Gahegan, 2012). Knowledge engineering is an actively evolving research domain within the field of artificial intelligence (AI), particularly symbolic AI, and recently in explaining and enhancing subsymbolic AI technologies (Forsythe, 1993; Lecue, 2020). The main objective of knowledge engineering is designing and constructing Knowledge-Based Systems that can use knowledge to deal with data in a similar way to humans (Studer et al., 1998). It involves acquiring domain knowledge from human experts, formalizing and encoding it using appropriate representation languages, and integrating it into systems as sketched in Figure 1-11. A key contribution of this PhD research is the formulation of a formal framework based on ontologies to capture and integrate expert knowledge into computational systems for 3D structural modeling, with particular emphasis on interpretative tasks and their associated knowledge. Consequently, this background section focuses on introducing the concepts of knowledge formalization and ontologies, while placing less emphasis on the other components of knowledge engineering, such as knowledge acquisition and complete system design.

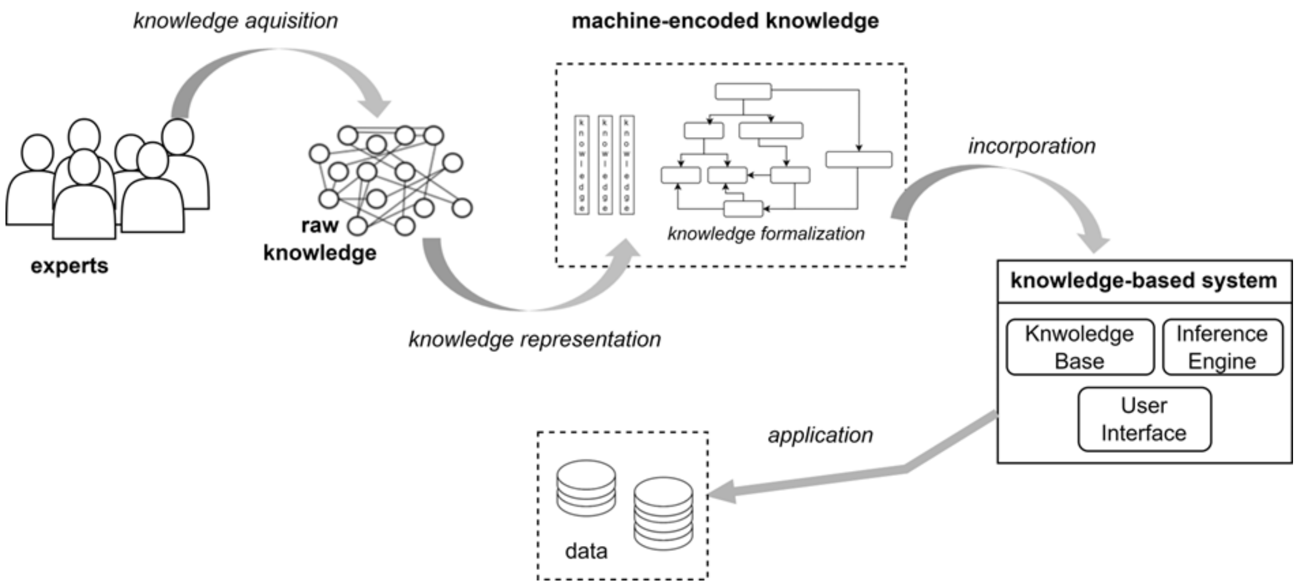


Figure 1-11 Sketch showing the steps for the creation of a knowledge-based system. Experts acquire raw knowledge from domain experts, which is encoded into a machine understandable language, including eventually formal languages. This is then incorporated into the system that contains this knowledge in a base, in addition to inference engine and user interface.

1.3.1 Knowledge formalization

Knowledge representation concentrates on representing knowledge in a way that a computer system can utilize for reasoning, problem-solving, and decision-making. Historically, representation included natural language descriptions, diagrams, semantic networks and logic-based languages (Lehmann, 1992; Newell et al., 1972; Newell and Tonge, 1960; Sowa and others, 1992). As the field progressed, challenges related to precision, interoperability, reasoning, and the systematic processing of knowledge led to the development of more *formal* approaches. Knowledge formalization, a subset of knowledge representation that seeks to express knowledge in an explicit, precise, and standardized format, often employing specific languages or systems. Description Logics (DLs) have become a prominent framework for this purpose, and ontologies used in this research are based on DLs.

1.3.1.a Description logics

Description Logics (DLs) are one of many approaches within the field of knowledge formalization aiming to describe knowledge by providing the mathematical and logical framework that supports formal reasoning (Baader et al., 2004; Nardi and Brachman, 2007). DLs dras upon concepts from both Set Theory and First-Order Logic.

- Sets are fundamental objects in mathematics, and set theory provides a formal framework for defining, manipulating, and studying them (Jiang et al., 2010). Sets can contain elements of any type, such as numbers, symbols, or even other sets.
- First-order logic –often referred to as predicate logic or quantificational logic– is a formal system used to represent statements about the relationships between objects in the world (Barwise, 1977). Statements are built using objects, predicates which represent properties or relations, and quantifiers, which indicate the scope of variables. Quantifiers include "for all" (∀), which denotes universal quantification, and "exists" (∃), which denotes existential quantification.

In DLs, sets represent collections of objects sharing common characteristics, while set operations denote relationships among these objects. DLs inherit the syntax of first-order logic, thereby establishing a formal language for expressing concepts and relationships within a domain. By integrating principles from set theory and first-order logic, description logics gain expressive capability in representing complex relationships and concepts within a domain. This combination facilitates the formalization of knowledge in a structured and systematic manner, enabling reasoning and inference mechanisms to operate effectively. From an engineering perspective, DLs use two notions (TBOX and ABOX) to separate the schema-level conceptualization of the domain from the instantiation of individual instances (Rudolph, 2011).

- TBOX (Terminology Box): This level defines the vocabulary and structure of a domain, including concepts. The TBOX is generally lamped with the RBOX that is about relations. For instance, in geology, concepts like "Deformation", "Process", and relation like "*has_Origin*" might be defined in the TBOX as: Deformation *has_Origin* Process.
- ABOX (Assertional Box): This level describes specific instances and their relationships within the domain. It asserts the existence of individuals (instances of concepts) and

relationships between them by applying the definitions from the TBOX to create statements about particular instances. For example, specific deformation structures (faulting1) with certain properties originate from a specific faulting episode (Miocene_faulting) would be described in the ABOX as: faulting1 *has_Origin* Miocene_faulting.

1.3.1.b Reasoning with description logic

Description Logics are not only used for representing but also reasoning about concepts, relations, and constraints in a domain (Baader and Nutt, 2003). Reasoning –also called inference– with the DLs is about making logical deductions or conclusions based on the information represented in a knowledge base applying logical rules, axioms, and constraints encoded within the formalized knowledge (see example in Figure 1-12). In the context of Description Logics, knowledge reasoning follows the principle of an open world assumption (Drummond and Shearer, 2006). This principle means that when reasoning about a concept, the system does not assume that any information not explicitly stated is necessarily false or does not belong to that concept. Instead, it acknowledges that the knowledge base may be incomplete, and that new information could be discovered later. For example, when a fault is described using DL, if this fault was not given its orientation parameters, reasoning will not consider this fault as inconsistent due to this lack, instead it assumes that this information could be later included. Therefore, the absence of specific information in an open-world scenario does not automatically imply a negative result, unlike in a closed-world assumption where the lack of information is often taken as evidence of non-existence. According to numerous authors (Baader et al., 2004; Baader and Nutt, 2003; De Giacomo et al., 1996; Rudolph, 2011) inference with the terminological level (TBOX) includes four main notions:

- Satisfiability: Determining whether a concept description is satisfiable with no inconsistencies.
- Subsumption: classification of concepts by determining whether one concept description is more specific than another concept description.
- Equivalence: Checking whether two concept descriptions are equivalent.
- Disjunction: Determining whether two concept descriptions are disjoint.

On the other hand, inference with the assertional level (ABOX) also includes four main notions:

- Consistency Checking: determining whether the ABOX assertions are consistent, i.e., whether they contain no contradictions with their definition in the TBOX.
- Instance Checking (Classification): determining whether a given individual belongs to a particular concept based on the asserted knowledge in the ABOX.
- Instance Retrieval (Realization): This problem involves retrieving all individuals that satisfy a given concept description. It is essentially the inverse of instance checking.
- Role checking: involves confirming whether specific roles between individuals are valid, inferring relationships and resolving inconsistencies between roles, determining the transitivity of roles, and verifying the existence of role chains to ensure a path of roles can be established based on asserted knowledge.

Reasoners are software tools that automate these reasoning tasks by taking as input a DL knowledge base and outputting new knowledge or answer queries. There are several popular reasoners available, each with its own strengths and areas of focus. For example HermiT (Glimm et al., 2014) and Pellet (Sirin et al., 2007) are de facto standards commonly employed by the community (Mishra and Kumar, 2011) due to their versality. As per Ponciano (2019) custom reasoners could be also developed based on the use case and the need using querying technologies combined with rule languages (section 1.3.2.b Building ontologies: representation, implementation and querying).

1.3.2 Ontologies

To formalize knowledge in information sciences, ontologies have emerged as powerful tools, and considerable endeavors have been devoted by researchers to their development. These tools are further employed in our approach to formalize expert's knowledge within a computable framework. From their origins in philosophical inquiry to their practical applications in information systems, ontologies have been interpreted in various ways across disciplines. One of the most widely adopted definitions in the context of information sciences comes from Gruber (Gruber, 1995) who defines as *[…an explicit specification of a conceptualization…]*.

This definition was improved by Studer et al. (1998) by including the concepts of formal and shared: *[…an ontology is a formal and explicit specification of a shared conceptualization…].* This definition highlights four key notions:

- Conceptualization: this term refers to the abstract representation of knowledge. It encompasses the ideas and theories that characterize the domain being modeled.
- Formal: this term implies a rigorous structuring using a formal language with unambiguous syntax and semantics using logical and mathematical formalisms. It ensures that the ontology can be processed by machines or humans.
- Explicit specification: means that an ontology shows the definition explicitly, on contrary to algorithms and mathematical methods.
- Shared: this term indicates that the knowledge defined in the ontology is a common understanding within a community or domain.

Ontologies are based on conceptual modeling, which is the process of creating abstract representations of systems, ideas, or phenomena. Conceptual models aim to capture essential features and relations, independently of how they will be implemented or represented in technology. In this sense, ontologies extend conceptual models by applying formal representation to these abstract structures. This formalization allows ontologies to explicitly represent the underlying structure and semantics of the knowledge in a domain.

1.3.2.a Components of an ontology

Ontologies are formalized using five types of components: classes, relations, axioms, rules, and instances. The Figure 1-12 presents a simplified example.

Class: refers to the representation in the ontology by type of material or an immaterial entity (e.g., Geological Unit, Fault, Shear, Faulting Process). According to Gómez-Pérez (Gómez-Pérez, 1999) the definition of a class should consider three aspects: the level of abstraction (concrete or abstract), atomicity (elementary or compound), and the level of reality (real or fictitious). To better consider the definition of hierarchies, and the variability of representations the work of Bachimont (2000) focuses on defining a semantic normalization according to the differential paradigm. This latter involves four key principles: community with the mother class, difference with the mother class, difference with the sister class, and community with the sister class. In their method OntoClean, developed for building and analyzing ontologies, Guarino and Welty (2009) suggests considering to describe a class the: rigidity (a class is considered rigid if all instances that belong to it necessarily do so in every possible context), identity (the ability to distinguish one individual from another within a set of individuals of the same class), unity (A class is said to have unity if it represents a single, coherent entity that is not merely a collection of smaller, distinct parts) and dependence (refers to the relational aspect where a concept's existence or definition is dependent on another concept).

Relations correspond to the type of possible interaction between the classes of a domain and their instances (e.g., a fault *results* from a faulting process). According to Bachimont (2000) a relation is defined by its signature and by the semantic content. For example, two relations can relate to two same concepts (equal semantic signature) without having the same meaning (different semantic content) (e.g., a class that references a knowledge model of geology could be related to the class person by the relation *has developer*, and by the relation *has user*). Two main types of relations can be distinguished: taxonomic relations and associative relations. The first, are used to hierarchically organize concepts (*is a*). The second, designates all relations between concepts which are not taxonomic relations. The associative relations can be further enriched by other properties such as symmetry, transitivity, and reflexivity.

Rules, in an ontology, are statements or assertions that define conditions to further structure the formal reasoning. These rules can be used to infer new insights or enforce certain logical conditions (see section DLs and reasoning). For example, if geological event ‘A’ is younger than geological event ‘B’, then the age of the object ‘a’ resulting from event ‘A’ must be less than the age of the object ‘b’ resulting from event B. They are often expressed using formal languages like SWRL (Semantic Web Rule Language).

Axiom is a constantly valid assertion about the entities (concepts and relationships). Axioms allow combining concepts, relationships and functions to define inference rules (e.g., Fault *has* Occurrence Time is an axiom). Cardinality constraints specify the number of instances of a property that can be associated with an instance of a class within an ontology.

Instances are used to represent the individuals of the class to which they belong (e.g., South Armoricain Shear is an instance of the class Shear).

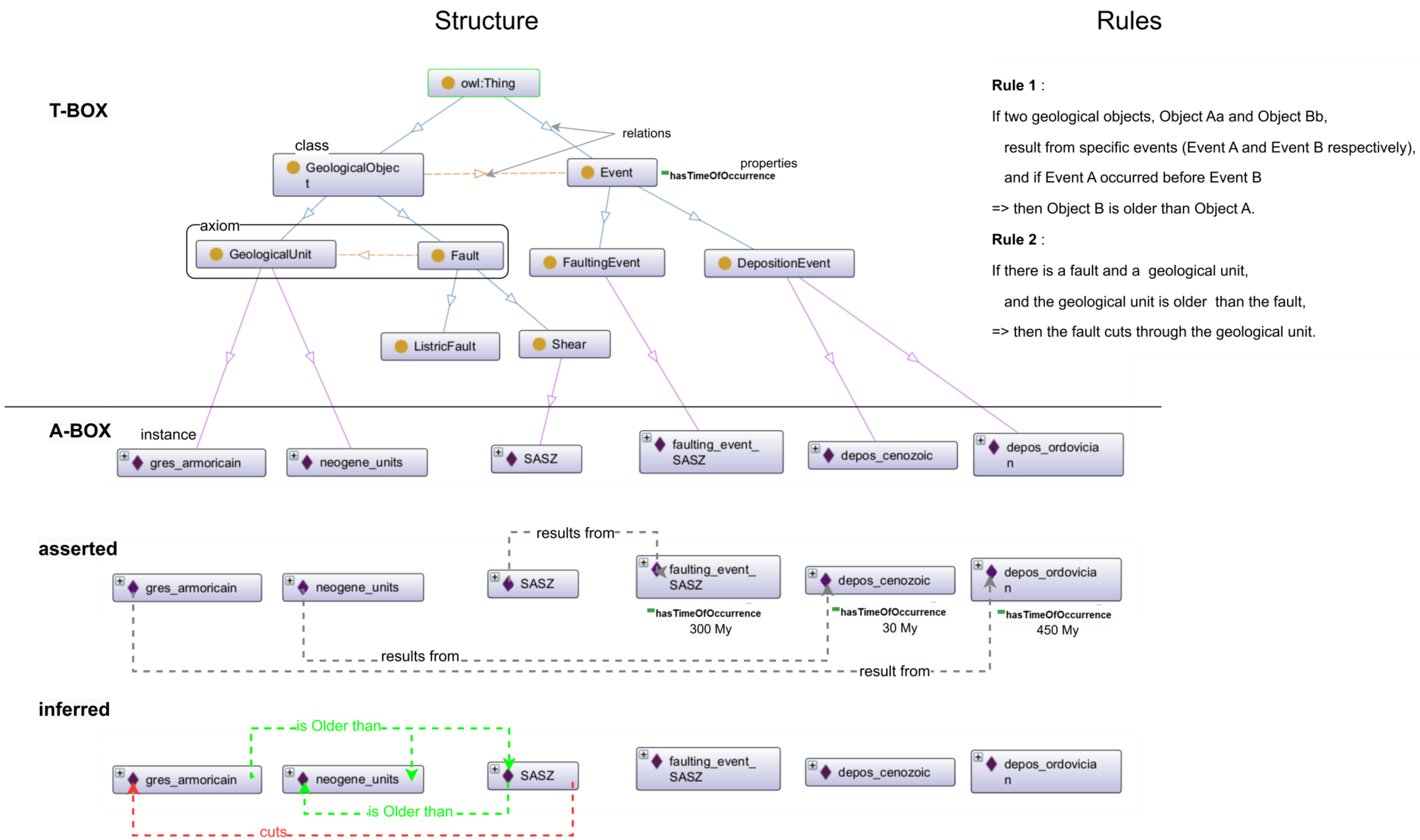


Figure 1-12 Example of a simplified ontology and its main components (see appendix B for link). The upper section (T-box) depicts the structure of the ontology, comprising 8 major classes, 5 relations, 1 property, and 2 rules. This ontology provides a framework for describing two types of geological objects (faults or geological units), the events that generated them, and the relative ages of these objects (e.g., older, younger) based on the precise timing of the events. The rules further constrain these relations. The upper section of the A-box illustrates the mapping of individuals to this ontology classes, including age assertions and their corresponding originating events. The lower section of the A-box shows the inferred relations generated from the assertions and rules introduced

1.3.2.b Building ontologies: representation, implementation and querying

Representation

Different languages exist for representing ontologies based on logical or non-logical paradigms like Ontolingua (Farquhar et al., 1997), Cycl (Fensel, 2000), LOOM (Corcho and Gómez-Pérez, 2000), Flogic (Frame Logic) (Fensel, 2000), and the Ontology Web Language (OWL) (Bechhofer et al., 2004). While the structure of these languages should not interfere with the way ontologies are developed, it's essential to note that language's design choices and constraints can influence how ontologies are conceptualized and structured. I think it is convenient here to cite the saying of the Austrian philosopher Ludwig Josef Johann Wittgenstein *[…the limits of my language are the limits of my world…]*. In matter of fact, these languages impose certain restrictions and conventions that developers need to adhere to, which might also limit the expressive power or flexibility in some cases. Moreover, the structure of these languages is tightly related to the reasoning capabilities which can influence ontology design decisions. I focus here on the Ontology Web Language (OWL) –employed in our approach– and I give some examples of its strength and weaknesses in term of expressivity and how this could influence the way ontologies are structured using it.

OWL represents one of many Semantic technologies that encompass various standards and protocols enabling the formal representation of knowledge in a machine-readable format as sketched in Figure 1-13. These technologies operate on the semantic level, focusing on the meaning and context embedded within the data, enabling machines to process data in a more aware manner (Berners-Lee and Swick, 2006). Thus, they play a crucial role in knowledge engineering and the semantic web which envisions the sharing of common and standardized knowledge and data across the internet (Berners-Lee et al., 2001). All Semantic technologies provide the technical infrastructure and standards for implementing and working with ontologies effectively within computer systems. However, OWL is the language for creating ontologies that provides a set of constructs for expressing entities. OWL facilitates the creation of hierarchies, and constraints allowing for the representation of complex knowledge structures in a formal and machine-readable format. OWL essentially implements the Description Logics –described above, designed to fulfill the needs of representing ontologies within the context of the Semantic Web.

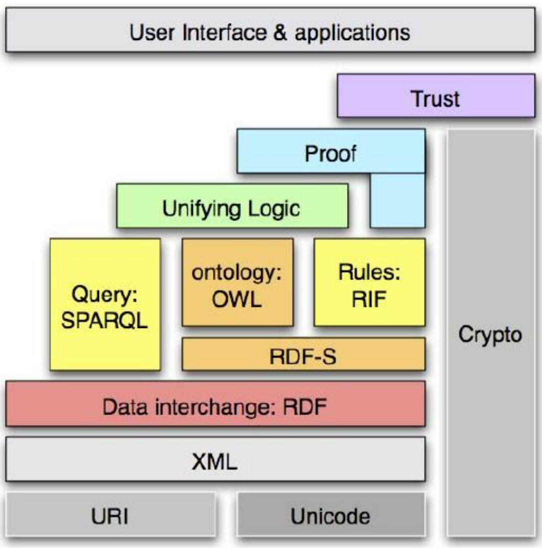


Figure 1-13 Semantic Web "layer cake" (Berners-Lee & Swick, 2006)

Semantic technologies are built upon each other. OWL is built upon URI, XML, RDF, and RDFs as illustrated in Figure 1-14.

- URI: stands for Uniform Resource Identifier. It is a string of characters used to uniquely identify a resource on the internet in a standard way (Jaffri et al., 2007). URIs can be used to identify various resources such as entities in an ontology (classes, relations), web pages, documents, and images.
- XML (eXtensible Markup Language): XML is a markup language used for encoding documents in a format that is both human-machine readable format (Wang et al., 2005). XML provides a way to structure data and metadata, making it easier to exchange and process information on the web.
- RDF (Resource Description Framework): RDF is a standard for representing information through triples usually used in the semantic web (Wang et al., 2005). A triple contains a subject-predicate-object statement. Subjects and objects are resources identified by URIs (Uniform Resource Identifiers). Predicates stand for relationships between these resources. Triplestores are databases designed to store and manage RDF data. RDF Schema (RDFS) extends RDF by providing a way to define vocabularies and describe relations between classes and properties. RDFS allows for a more structured and semantically rich representation of data on the web. With RDFS, it is possible to define hierarchies of classes and properties, specify domain and range constraints, and establish relationships between classes and properties.

Expressivity

Unlike RDFS, which primarily focuses on defining basic taxonomies and relations, OWL introduces more advanced mechanisms for expressing complex conceptual relations. One key aspect is its support for class comparison mechanisms, which enable the precise specification of relations between classes. For instance, OWL allows the definition of equivalence between classes, indicating that two classes have the same instances. It also supports concepts like symmetry, transitivity, and reflexivity, providing a sophisticated framework for reasoning about classes and their relationships. These features enhance the expressivity of OWL.

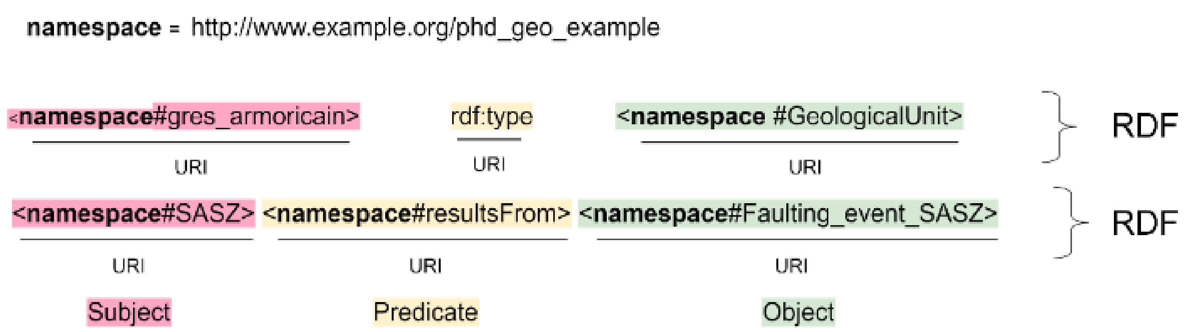


Figure 1-14 RDF representation of geological objects used in the example above (components of an ontology) using the defined namespace http://www.example.org/phd_geo_example. The first RDF triple defines gres_armoricain as an instance of the class GeologicalUnit, with rdf:type as the predicate. The second RDF triple links the geological faulting event SASZ to the faulting process that generated it using the resultsFrom predicate

While OWL offers significant advantages for representing knowledge statically, it has limitations when it comes to dynamic representation of this knowledge. One key weakness is its inability to directly represent temporal dynamics or state changes over time that implicates reclassifications of instances. For example, a sedimentary rock (instance of the type sedimentary rock such as shale) transforms into a metamorphic rock under heat and pressure (becomes instance of a metamorphic rock class slate). This process involves a state change over time, but OWL cannot natively handle the dynamic reclassification of the instance rock from "Shale" to "Slate" as conditions evolve.

Another limitation is the inability to model ternary relations or relations involving more than two entities, which frequently arise in dynamic scenarios where two entities interact in the context of a third (e.g., a specific deformation style occurs on a specific rock type in a specific geological setting). Moreover, OWL lacks native support for relations between relations, also known as higher-order relations. These are essential for representing complex dynamics where relations themselves evolve or interact over time. For instance, if a geological unit undergoes contraction, it subsequently enters a deformation process. The relations of undergoing contraction and entering deformation cannot be directly linked in OWL.

Many of these limitations can be addressed through the use of rules; for example, if the relation of undergoing contraction is asserted, the relation of entering deformation can also be inferred. However, OWL does not natively support rules, which has led to the development of extensions such as the Semantic Web Rule Language (SWRL) by the community (Horrocks et al., 2004). SWRL is a rule-based language for the Semantic Web that combines OWL (Web Ontology Language) and RuleML (Rule Markup Language). It allows defining rules in the form of if-then conditions. In addition to SWRL, other rule languages such as SHACL (Shapes Constraint Language) (Pareti and Konstantinidis, 2022) are also available for inference and validation tasks within ontologies. These rule languages provide alternative approaches for expressing constraints and inference rules over RDF data and ontologies (Rattanasawad et al., 2013). However, it is important to note that while these rule languages offer flexibility and customization options, caution must be exercised when using them, as they may not be compatible with standard OWL reasoners.

Implementation

After the conceptualization and formalization of a given ontology, a large choice of editing tools and software are available for its implementation. In this research we use Protégé[5] (Noy et al., 2003) and Owleady2[6] (Lamy, 2021).

- Protégé: Protégé is one of the most widely used ontology free editing tools. It provides a user-friendly interface for creating, editing, and visualizing ontologies. Protégé supports multiple ontology syntaxes and offers various features such as ontology visualization, reasoning support, and plugin extensibility.

[5] https://protege.stanford.edu/

[6] https://owlready2.readthedocs.io/en/v0.47/

- OwlReady2, is a Python library for ontology-oriented programming. it enables developers to create, manipulate, and reason over ontologies in a pythonic programming framework.

An ontology could be serialized in different formats like RDF/XML, Turtle, and JSON-LD and in different syntaxes like the Manchester syntax (Staab and Studer, 2009).

Querying ontologies: SPARQL

OWL ontologies could be queried using different tools and languages such as SeRQL (Ekaputra et al., 2013), RDQL (De Laborda and Conrad, 2005), OWL-QL (Fikes et al., 2004), OntoQL (Jean, 2007), and SPARQL (Pérez et al., 2009). SPARQL (Semantic Protocol and RDF Query Language) is the most common querying language and is adopted by the W3C (World Wide Web Consortium). It allows users to retrieve and manipulate data stored in RDF format, enabling complex queries over graph-structured data. SPARQL supports various querying operations:

- SELECT to retrieve specific variables (e.g., all instances of a class, specific properties) from the RDF database.
- CONSTRUCT to create or transform RDF graphs (e.g., results of inference) based on a query.
- ASK to check if an RDF graph corresponds to a specific query pattern (e.g., to verify the existence of certain conditions in the knowledge base).
- DESCRIBE to get a description of specific RDF graph pattern.

1.3.2.c Building ontologies: methods and practices

In his paper titled "Understanding, building and using ontologies" (1997), Guarino provides comprehensive guidance on best practices for constructing ontologies:

- The need for a thorough analysis of the domain to capture the essence of the modeled domain.
- Building ontologies with keeping in mind that knowledge can be shared and applied across different applications, reusability is a crucial aspect of ontologies.
- The use of different levels of ontologies as it allows for a more modular and adaptable approach to knowledge representation.
- The use of formal properties like rigidity, identity, unity and dependence to clarify the meaning and relationships between concepts within the ontology.

The life cycle of an ontology could be classified into five main stages (Pinto and Martins, 2004):

- Specification: Determining both its purpose and scope.
- Conceptualization: Developing a conceptual model of existing entities that aligns with the project specifications.
- Formalization: involves translating the conceptual description into a formal model.
- Implementation: Implement the formalized ontology in a representation language.
- Maintenance: Updating and correcting the implemented ontology.

Ontologies are developed through a structured process. This latter could be straightforward or more complex depending on different factors such as domain of application, the level of detail required, the availability of existing resources, and the expertise of the developers. Main methodologies of building ontologies according to Fernández-López and Gómez-Pérez (2002) are: Cyc, Uschold and King, Grüninger and Fox's, KACTUS, METHONTOLOGY, and SENSUS methodologies. These methodologies share the focus on emphasizing stages, consideration of ontology documentation, the recognition of the need for balance between detail and categorization, and the need for defining explicit competency questions that the ontology will address (Monfardini et al., 2023; Sousa et al., 2014). However, they differ in the approach to concept identification, level of formality, inclusion of re-engineering guidelines, scope and domain specificity, and the depth of documentation provided. These methodologies are centered on two prevailing paradigms: Top-down and Bottom-up.

- The Top-down paradigm describes the starting by general abstract concepts and going down to more specific domain concepts.
- The Bottom-up paradigm describes the beginning with instances derived from existing data, then extracting concepts, relationships, and constraints and organizing them into an ontology structure.

None of these methodologies is a worldwide standard for all applications (Fernández-López and Gómez-Pérez, 2002). Any of them could be used for developing an ontology and coupling two methods or more is also possible depending on the case study (Ra et al., 2012).

1.3.2.d Typology of ontologies

Ontologies can be classified based on different criteria such as purpose, formalism, structure, granularity and abstraction. Using abstraction, ontologies developed across different domains are layered into four levels (Haller and Polleres, 2020). For reference, some examples of each category are cited, but their detailed description is out of the scope of this PhD. Cited ontologies used in this work will be thoroughly detailed in corresponding sections of usage:

- Upper ontologies, top-level ontologies or foundational ontologies: define general terms common across all knowledge domains. They define basic concepts such as space, time, objects, and their relations. Examples of which are: the CYC (Foxvog, 2010), Suggested UMO (Niles and Pease, 2003), DOLCE (Borgo et al., 2022), and BFO (Otte et al., 2022).
- Mid-level ontologies or global domain area ontologies: act as bridges between abstract content of upper ontologies and the richer detail of domain ontologies. Possible examples of mid-level ontologies are: PROV-O (Lebo et al., 2013), FOAF (Kalemi and Martiri, 2011), SSN (Compton et al., 2012), and the IAO (Ceusters and Smith, 2015).
- Domain ontologies capture entities relevant to a specific domain of knowledge either with a core set of essential concepts or through a more exhaustive representation. Examples include: the Gene Ontology (Ashburner et al., 2000) or OBI (Bandrowski et al., 2016), GeoCore (Garcia et al., 2020), and GSO (Brodaric and Richard, 2020).
- Use case ontologies are tailored to support specific tasks or applications with detailed classes and relations highly dependent on the use case. For example, the GeoFault ontology used for interoperability of fault data in 3D modeling (Qu et al., 2023).

1.4.Formal knowledge for 3D structural modeling

In recent years, knowledge models based on semantic technologies, particularly ontologies, are increasingly developed and adopted by the geoscientific community. To provide a clearer understanding of what this research offers compared to existing solutions, this part will briefly review the current state of science regarding main applications and advances of formal knowledge in the field of 3D modeling, and geosciences in general.

1.4.1 Formal knowledge in 3D structural modeling

3D structural modeling field involves representing geological entities. Therefore, formalizing knowledge for this field is fundamentally related to formally describing these entities and all the processes involved in the creation of these geometric representations. As discussed next, existing literature focuses more on using ontologies –or in general semantic models– to formalize knowledge related to the description of geological entities, as most developments have been driven by the need for data sharing, storage, and querying. However, these applications have paid less attention to the interpretation and in general the modeling processes that lead to building these models. Moreover, less attention is paid to the uncertain fallibilistic aspects that characterize the geoscientific information.

Recent of works of Zhan et al., (2022, 2019) focus on describing already constructed 3D structural models through their mathematical properties in a formal way to capture the geometric and the entangled geological semantics of these models based on the notion proposed in the work of Thiele (2016a) (see Section Structural modeling). The work of Zhan et al., is based on mapping the geological semantics to an existing ontology "Structural Geology Ontology" (Babaie et al., 2006). Endeavors such as those of Cicconeto et al. (2022) and Qu et al. (2023) concentrate on defining more specialized ontologies for reservoir modeling in the petroleum industry: for marine sedimentary deposits and their characteristics (Cicconeto), and describing faults, their components and characteristics (Qu).

On a more general level of sharing data in the 3D modeling field, a pivotal contribution is the RESQML standard (Reservoir Model Markup Language), detailed by Morandini et al. (2011). RESQML provides a standardized XML-based format for exchanging subsurface information, including geological and engineering aspects of reservoir exploration and exploitation in the petroleum industry, in addition to addressing some notions of interpretation and representation in relation to 3D models. However, despite its focus on ensuring information exchange between systems in a semantic framework through URIs and schemas, RESQML is not an ontology per se as it does not focus on the conceptual modeling and reasoning aspects required in constructing ontologies and the definitions of entities.

In relation to the RESQML, contributions from Belaid et al. (2011) and Mastella (2009) focus on providing a semantic framework based on the indexation and annotation of modeling workflows (e.g., seismic surveys, seismic processing) for Belaid, and the heterogeneous data generated by various exploration and exploitation activities (e.g., seismic surveys, well drilling, well logs) for Mastella.

1.4.2 Formal knowledge in geosciences

Various other ontology-based applications have emerged across different fields to describe geological entities over the last two decades, offering diverse methods for formalization according to specific needs. For instance, Zhong et al.(2009). provide an ontology for fractures, Lombardo et al., (2018) work on capturing the knowledge in geological maps in an ontological framework based on the GeoSciML (Simons et al., 2006). The GeosciML (Geoscience Markup Language) is another XML-based standard that is developed to encode and exchange data in geological mapping. The work of Lombardo is extended by Mantovani et al., (2020) to include geological events as entities in the encoding of geological maps.

Although less related to 3D modeling, other notable works in the use of ontologies to represent geological entities include Babaie et al. (2020) who for example apply ontologies to microscope analysis. Abel et al. (2012a) use ontologies for facies descriptions to characterize lithologic logs, while Abel et al. (2019) propose an approach to organize visual content in the petroleum industry using a Knowledge-based system for assisting machine learning algorithms. Fauziati and Watanabe (2010) conducts a research on using ontologies to help in hazard assessment in volcanic regions. Wang et al., (2018) extend the work of Ma about the use of ontologies in managing time-scale information (2012) to include paleontological information.

1.4.3 Gaps in Knowledge Formalization for 3D Structural Modeling

According to Whitehead and Gahegan (2012), developed ontologies in geosciences are typically designed for single-purpose use. Although this analysis was established in 2012, it remains relevant to a certain extent, as many of these ontologies continue to be isolated for their specific applications. Most of these ontologies lack a common framework for describing geological entities that are shared across different applications. This fragmented approach of developing ontologies is mainly coming from the absence –until recently– of domain ontologies in geosciences that could provide a unified framework for describing geological entities. An exception to this absence is the SWEET initiative (Raskin and Pan, 2005). SWEET (Semantic Web for Earth and Environmental Terminology) is a general ontology developed by NASA* that covers a broad range of concepts related to Earth and environmental sciences. However, it has a low level of ontological commitment and emphasizes general Earth and natural processes over specialized geologic concept.

To address the rising need for a common framework, two main ontologies have been proposed to describe geological entities: GSO (Brodaric and Richard, 2020), and GeoCore (Garcia et al., 2020). In contrast to SWEET, both GSO and GeoCore provide more ontological commitment and specialization in geosciences. These two are further detailed in Chapter 3 to avoid redundancy. But it is important to mention here that although these two domain ontologies –GSO, GeoCore– have some fundamental differences in their definitions of geological entities, they both provide a solid framework to build more specific applications on top of them, thereby ensuring consistency in definitions and interoperability between systems. Withal, these ontologies are not fully adopted by the geoscientific community as they are recently developed, therefore, their application is still limited to the developing groups.

*https://www.nasa.gov/

Whitehead and Gahegan (2012) also emphasize that it is insufficient to use semantics merely for assigning categorical labels to geological entities. Given the interpretive and uncertain nature of geosciences, researchers need to understand how conclusions are reached to confidently adopt them. Therefore, the use of ontologies ought to include the epistemological questions inherent in the interpreted geological information and the geocognitive processes engaged. This analysis is extremely pertinent when addressing the formalization of knowledge for 3D geomodelling, where epistemic uncertainty is heightened due to limited data and the need to extrapolate knowledge from narrow segments of known reality. Capturing or providing a framework for these epistemological aspects is challenging, as it involves multiple components, including the interpretation, cognitive processes, epistemology, and the processes of knowledge creation and learning. From a formal perspective, in the geoscientific field, including 3D structural modeling, this area of literature is significantly underdeveloped, with only a few works addressing this issue directly. For instance, Brodaric (2012) focuses on the ontological modeling of inference processes employed in creating knowledge in geological mapping, while Carbonera et al. (2013) explore the ontological modeling of inferential knowledge used to interpret visual content through the development of a rule based system based on the conceptual generation of hypotheses. Other works address these issues indirectly by developing ontologies that provide rule-based systems to assist in constructing 3D models as in the work of Perrin et al. (2005) as discussed further, thus providing a certain record of modeling tasks. But in general, these ontologies do not model the interpretation itself and related process as separate entities, and do not associate properties of uncertainty or evaluation and related decisions to geoscientific information.

1.5. Summary & methodological discussion

The first section of the current chapter delves into the fundamental geological elements that 3D modeling aims to capture with a particular focus on their structural characteristics. It also introduces the concept of geocognition, highlighting the distinctive mental abilities that underlie geological thinking. In addition, it provides an overview of the current state of 3D structural modeling, examining various methodologies, representation techniques, and the model-building process. It highlights the role of the judgment of experts in compensating for the limitations of modeling tools that often struggle to cope with the natural geocognitive abilities of experts and how they use knowledge, which lead often to more limitations in knowledge sharing, communication, and application. Current methods and workflows that assist in the interpretation of geological entities and their characterization, propose only rigid templates to perform certain tasks often based on predefined frameworks from a mathematical perspective, which are difficult to update or modify. These templates implement strategies that work well in similar cases or modeling environments but do not in any way address the interpretation as a distinct process that needs to be formalized by itself. The second section of this chapter introduces the fundamental concepts related to knowledge formalization with a focus on ontologies. It outlines the key technical principles and guidelines for developing, using and manipulating ontologies. Additionally, it provides an overview of the current research on formalizing expert knowledge in the field of 3D structural modeling and geosciences in general, identifying the existing gaps that this research work aims to address.

As outlined by the previous discussions, there is a clear contrast between the knowledge utilized by experts and the capabilities provided by current geomodelling systems to interpret and use this knowledge. This contrast necessitates an alternative approach that extends beyond treating the problem merely as one of interpolation and mathematical knowledge incorporation. This alternative approach must consider the interpretive aspect that enables systems to interact with interpreters at a more 'geological' level. This perspective inherently involves considering the conceptual understanding that underlies the modeling process, along with deeper knowledge embedding and consideration of epistemological aspects associated with geoscientific information, and eventually how this information is modeled and represented.

These key elements are fundamental for advancing the automation of the interpretation process and improving the interaction and knowledge sharing amongst experts and geomodelling systems. Thus, this research will address these challenges by positioning itself at the intersection of multiple disciplines –geology, 3D modeling, cognitive sciences, conceptual modeling and knowledge formalization. A methodological approach that holistically integrates these diverse aspects will be developed and refined over the course of this work.

Adopting an interpretive perspective requires formalizing the interpretation process itself as a distinct entity, rather than only modeling its outcomes. This formalization is inherently related to the use of structured knowledge that encapsulates the comprehensive understanding of experts. As proposed in this dissertation (Chapter 2), it is feasible to establish a simplified knowledge framework that provides the essential components necessary for interpretation while focusing on a generic framework that describes how experts in 3D structural modeling engage in interpretation. The fundamental components of this knowledge framework include geological, interpretative, and representational knowledge aspects, which can be further tackled from a more formal point of view, as done in this PhD research.

In fact, to effectively apply the proposed approach in more complex case studies, where knowledge is formally described, the previously identified knowledge aspects (geological, interpretative, and representational) require additional formalization and conceptual modeling. This includes examining the relations between models, representations, and the frameworks representing processes involved in interpretation and modeling. These aspects are thoroughly explored in Chapter 3, allowing further capture of knowledge applied tacitly by experts in classical modeling approaches. Knowledge formalization is an emerging discipline within geosciences. Therefore, the need is for reliable knowledge models that not only address the specific requirements of the 3D modeling field and its applications, as well as the geological field but also align with the domains of knowledge engineering and formalization, enabling reusability and interoperability with other geoscience applications. In response to these needs, the second stage of this research (Chapter 3) places a strong emphasis on knowledge formalization and conceptual modeling, with an ontological commitment guiding the development of the proposed knowledge model.

Chapter 2: Formalization of the interpretation process

Contents

Abstract

This chapter introduces the formalization of the interpretation process, with a focus on leveraging geological knowledge explicitly within a structured framework to construct 3D structural models. To demonstrate this formalization, a prototype algorithm is presented as a proof of concept. The algorithm processes inputs (e.g., field observations), identifies areas with incomplete information or ambiguity requiring interpretation, and performs the interpretation through a continuous improvement cycle (Plan, Do, Check, Act). While the algorithm currently operates in a semi-automatic manner, it marks a significant step toward achieving full automation in the future.

The content of this chapter is the equivalent of the paper (A knowledge-driven modeling formalism for automatic structural interpretation) published in the Earth Science Informatics journal under the DOI: https://doi.org/10.1007/s12145-024-01613-y and is authored by: **Imadeddine LAOUICI, Gautier LAURENT, Christelle LOISELET, and Yannick BRANQUET**. It is integrated herein as is in its post-print version with additional insights and discussions.

2.1. Introduction

Three-dimensional (3D) structural geological modelling aims to produce numerical versions of the geological architecture of some part of the Earth's crust (Perrin and Rainaud, 2013). Resulting structural models provide a comprehensive understanding of the geometries of geological objects, including beds, folds, and faults, along with their relationships (Laurent, 2013; Laurent et al., 2016; Patel et al., 2021; Thiele et al., 2016a). Structural modeling finds extensive applications across various fields of geoscience such as resource exploitation, geological engineering, natural hazard mitigation, and groundwater management. This paper introduces a paradigm to build 3D structural models in a semiautomatic manner, building on related work that uses formalized knowledge for structural modeling (Perrin et al., 2005), with "formalized" meaning the use of an explicit formalism to build sharable knowledge between users and systems (Abel et al., 2012b). Our approach advances existing work by formalizing the interpretation process typically carried out by geology experts.

In 3D structural models, geological objects are often represented via structural surfaces (e.g., tops and bottoms of geological units, and faults) (Perrin and Rainaud, 2013). The various modeling algorithms (Section 2.2) mainly differ in how they represent the surfaces and compute the values of their properties (Guo et al., 2020; Patel et al., 2021; Wellmann and Caumon, 2018). While these algorithms excel at constructing geometries and topologies (Wellmann and Caumon, 2018) they are heavily reliant on an expert's involvement and have limited ability to aid in interpretation. Interpretation, in this context, refers to the process of analyzing and explaining input data (e.g., field observations, geophysical data) with respect to existing knowledge, such as known geological laws and principles (Bond et al., 2011, 2015; Frodeman, 1995, 2014). As a by-product, the interpretation process also infers objects in blank areas where inputs are missing and offers explanations for these areas. Consequently, during the construction of the geomodel, most interpretive tasks are left to experts: e.g. selecting appropriate inputs, determining which inputs to consider at each step, and selecting the interpolation algorithm as well as configuring its parameters. Additionally, the identification of geological objects, such as suggesting folds to explain a surface dip variation, remains primarily an expert task, so decision-making in existing approaches is largely reliant on expert judgment.

Another limitation of many modeling algorithms is their reliance on geometric representations of geological objects, as surfaces or vectors, instead of consideration of the objects in their full real form, e.g. as volumetric entities with a wealth of geological characteristics. They thus take, in effect, a representational versus ontological approach to real objects. Consequently, some aspects of the modeled objects that are easily conceptualized by experts are excluded from the modeling tools if they are difficult to represent spatially or mathematically. For example, something as essential as a layer's thickness, which is used to interpret the geometry of a unit, is hard to incorporate in an interpolation framework (Laurent, 2016), as thickness itself does not occupy space, being an immaterial characteristic like size or shape, nor is it inherently a vector, though it might be derived from various vectors with some effort. As another example, faults are often represented by a simple single surface (Caumon et al., 2004; Mallet et al., 1989), even though other features (e.g., fault core, multiple segments) are used to interpret fault geometry.

In the end, only limited geological knowledge is typically incorporated into geomodelling tools and resulting models (Kessler et al., 2009; Kolesov, 2018; Royse, 2010), and large parts related to interpretation remains resident in expert's minds. In addition, only a limited number of existing numerical approaches use formalized knowledge to aid interpretation. For example, Perrin et al. (2005) propose an approach to automatically crosscut structural surfaces using a formal rule-based framework that describes the spatiotemporal relations of the surfaces. Although the approach does include various properties of geological objects, surfaces must be first constructed by experts through interpolation, thus a significant part of the interpretation process is reliant on expert input.

As expert input is often subjective, and the results are not easily explained nor reproduced, it would be beneficial to have a system that could not only assist in the interpretation process, by making suggestions and validating against background knowledge, but also recording the knowledge used, inferences made, and steps taken. Such a system would minimally clarify subjective assumptions and ideally advance reproducibility and explanation. From our perspective, the main obstacle to such a system's development is the lack of formalization of the interpretation process. To address this obstacle, the current chapter focuses on the following questions:

- What are the main tasks (e.g., selection of study location and inputs) performed during the interpretation process? How to formalize these aspects in a generic and algorithmic way?
- As the formal interpretation process will require a knowledge framework, which knowledge aspects are needed and how to structure them?
- How to include the diversity of properties of geological objects in the knowledge framework, and how to describe them beyond their mathematical formulation?

The current chapter is structured as follows: Section 2 describes a synthetic modelling scenario used to illustrate the algorithm and its implementation; Section 3 introduces theoretical aspects of the proposed approach that are implemented in a prototype algorithm in Section 4; this prototype implementation highlights some interesting considerations discussed in Section 5; and the paper concludes with a brief summary in Section 6.

2.2. Synthetic scenario of structural modeling

This section introduces a modeling scenario –a use-case– from synthetic inputs. The main objective is to exemplify the modeling and interpretation process in a generic and explicit manner, demonstrating how experts approach it, and what existing solutions offer. In doing so, existing limitations related to interpretation and expert involvement are highlighted and these are addressed in the remainder of the chapter.

2.2.1 Problem statement

The Figure 2-1 shows a simplistic dataset – in a vertical cross-section – that could be used to build a structural model representing a deformed geological interface (surface_1). The dataset contains 11 data points having spatial and geological properties (e.g. geospatial location, dips, and polarities). This scenario raises further questions that need to be resolved, in addition to the general questions outlined in the introduction:

- Type identification: which types of geological objects are present?
- Property identification: what are the spatial and other properties of the geological objects?
- Spatial continuity: which input properties are continuous enough in space to overcome the scarcity of data?
- Superposition: what are the relative temporal and topological relationships between the identified objects?

Resolution of these questions is affected by the nature of the data inputs (point-based), as well as their density (11 points) and distribution (heterogeneously placed in a 3D space).

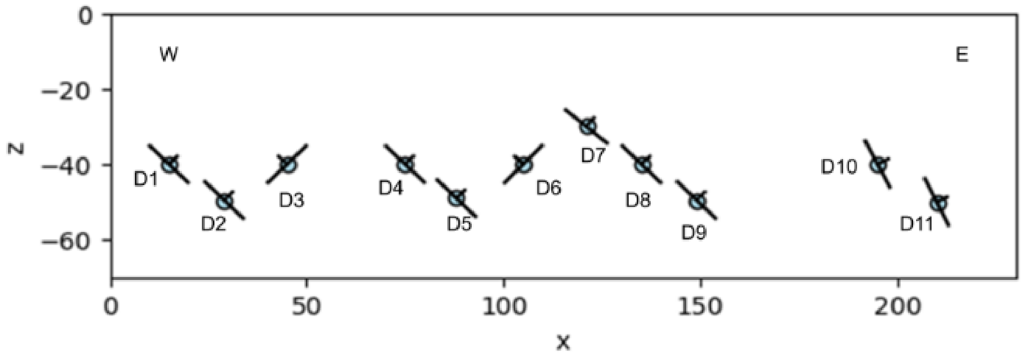


Figure 2-1 Vertical West-East section showing a dataset with 11 point-based observations. Light blue disks denote observations; long black lines denote bedding plane apparent dips in the cross-section plane; and short black lines show younging direction. The figure is scaled in meters along the x, y, and z axes, though the approach remains scale-neutral. Observation points are distributed in 3D space and projected onto the cross-section for visualization purposes.

2.2.2 State of the art: Expert-based process

3D structural models are often constructed using interface-based methods, where structural surfaces –e.g. top and bottom surfaces as well as fault planes– are represented and managed based on geological knowledge (Caumon et al., 2009; Mei, 2014). According to Wellmann and Caumon (2018), two approaches are used by experts to generate interfaces: explicit and implicit. Explicit approaches represent interfaces as geometrical objects, typically as triangulated surfaces (Mallet, 1992) or parametric surfaces (De Kemp, 1999), possibly developed from instance depth functions (Hardy, 1971). In contrast, implicit approaches derive surfaces from isovalues of a scalar field defined by an implicit function (Calcagno et al., 2008; Caumon et al., 2013; Lajaunie et al., 1997). Explicit representations offer more control on local geometry and topology (Collon et al., 2016); conversely, implicit representations facilitate the modeling of multiple conformable layers and their update (Calcagno et al., 2008).

To deal with the challenges outlined by the use-case, contemporary solutions provided by geomodelling tools heavily rely on interpolation to estimate properties of structural surfaces – in both explicit and implicit representations– in areas with sparse or incomplete data, with many significant advancements (Wellmann and Caumon, 2018). Numerous interpolation techniques have been developed to generate values of properties of geological objects based on data distribution (De Kemp, 1999, 2021; Lyu et al., 2021; Mallet, 1997; Sprague and De Kemp, 2005; Watson, 1999; Webster and Oliver, 2007).

While a comprehensive review of all these techniques is beyond the scope of this paper, it is important to note that each technique has its own advantages and disadvantages (Wellmann and Caumon, 2018). For instance, Kriging is a geostatistical method that accounts for spatial correlations and variability between input data on either a global or local scale (Fabbri et al., 2020; Lajaunie et al., 1997; Renard et al., 2019). This makes it particularly effective for capturing complex spatial relationships and providing unbiased estimates with minimized variance based on the variogram model. In contrast, Radial Basis Functions (RBF) (Cowan et al., 2002; Hillier et al., 2014) focus on localized influences, excelling at capturing smooth and continuous variations, especially within a well-defined local neighborhood. It is important to mention that although the underlying mathematical assumptions and mechanisms of both techniques differ, when applied in a global context, both Kriging and RBF can yield similar interpolated distributions.

Despite the great potency of tools based on these methods, these approaches inherently lack the capability to interpret like experts, requiring substantial user involvement to select pertinent inputs, identify objects, and infer structures (Wellmann and Caumon, 2018). Recent advancements, such as Laurent et al. (2016) and Pizzella et al. (2022), integrate geological knowledge to enhance characterization of structures like folds. However, the identification and validation of such deformation objects still relies on user input, making interpretation predominantly a human-driven process. Moreover, solutions that do formalize parts of the interpretation process are narrowly focused, e.g. concentrating on fault detection (Di and Gao, 2016; Godefroy et al., 2019) or estimating fold characteristics (Grose et al., 2017, 2018).

Even geometric representations built using tools integrating some knowledge aspects are not always valid geologically (Zhong et al., 2020). For example, interfaces generated by explicit methods often require expert involvement to adjust their geometry and topology (e.g., trimming, smoothing) (Caumon et al., 2009) and to establish relations between them (e.g., cross cutting) (Caumon et al., 2004). Output of implicit methods also often require expert involvement to validate the geometry of complex structures (e.g., with closed surfaces) and to define conformable/non-conformable sequences (Belhachmi et al., 2024). Efforts to incorporate more knowledge into modelling, such as Laurent (2013, 2016), Laurent et al, (2016), and Grose et al. (2019, 2021), are constrained by the interpolation-centric frameworks; e.g. constraints on thickness are difficult to consider in a linear system of equations without adding an iterative process (Laurent 2016) or an inverse framework (De La Varga and Wellmann, 2016). Moreover, although Wellmann et al. (2014), Olierook et al. (2021) establish a framework for validating models from geological knowledge, the focus is on the mathematical encoding of the knowledge and still leaves the selection of the knowledge to experts (Wellmann et al., 2014b).

Given the involvement of experts remains essential, regardless of the chosen approach, we expand our use-case on a step-by-step basis to illustrate the dynamics of traditional 3D modeling. Tools then provide interpolation solutions, and experts leverage their knowledge, adapting it to fit the modeling tool (Figure 2-2). Note this use-case is merely illustrative and not a rigid template: different experts might conduct these steps in varying ways. The key stages include the following, and are roughly aligned with Brodaric (2012):

- **Hypothesis generation**: at the outset, the expert modeler runs an initial interpolation (in this case via RBF) on all available inputs to visualize a preliminary construction of the presumed geological interface, surface_1, and explore its general behavior over the studied area. Results suggest the presence of some deformation pattern (upper part of Figure 2-2). The observed deformation pattern could be interpreted by experts as one or more folding structures, or some faulted displaced blocks. The generation of such hypotheses is intuitive for human interpreters familiar with deformation concepts, but challenging for machines (Oakley et al., 2024). For sake of example, let the chosen hypothesis be the existence of some folds.

- **Object individuation**: once the hypothesis is chosen, individual objects are identified and characterized, e.g. their geometry in terms of wavelength and amplitude. Additionally, relations between the identified objects are determined, e.g., shared limbs, how folds are related to surface_1, as shown in the middle part of Figure 2-2. Based on the continuity principle (Rosenberg, 2001) and in the absence of other specific folding patterns, regular folding seems to be a plausible assumption (Khaninezhad et al., 2012) as it satisfies the data and requires less parameterization compared to more complex geometries.

- **Object characterizatio**n: following the choice of regular folds, experts iteratively introduce additional constraints, which is a subjective task involving decisions on both the location and quantity of geometric and other characteristics of the constructed interface, as sketched in the middle part of Figure 2-2. Experts ensure the validity of the results in each iteration, e.g., ensuring the coverage across the entire study area and preventing issues caused by interpolation (surfaces mismatch or artefacts).

- **Model representation:** the final model is represented by the interpolation results as sketched in the lower part of Figure 2-2, while interpretation decisions remain as personal insights of the modeler.

At its end, although experts use interpolation in all steps, what is most critical is their understanding and interpretation of the objects and relations that guide the interpolation itself.

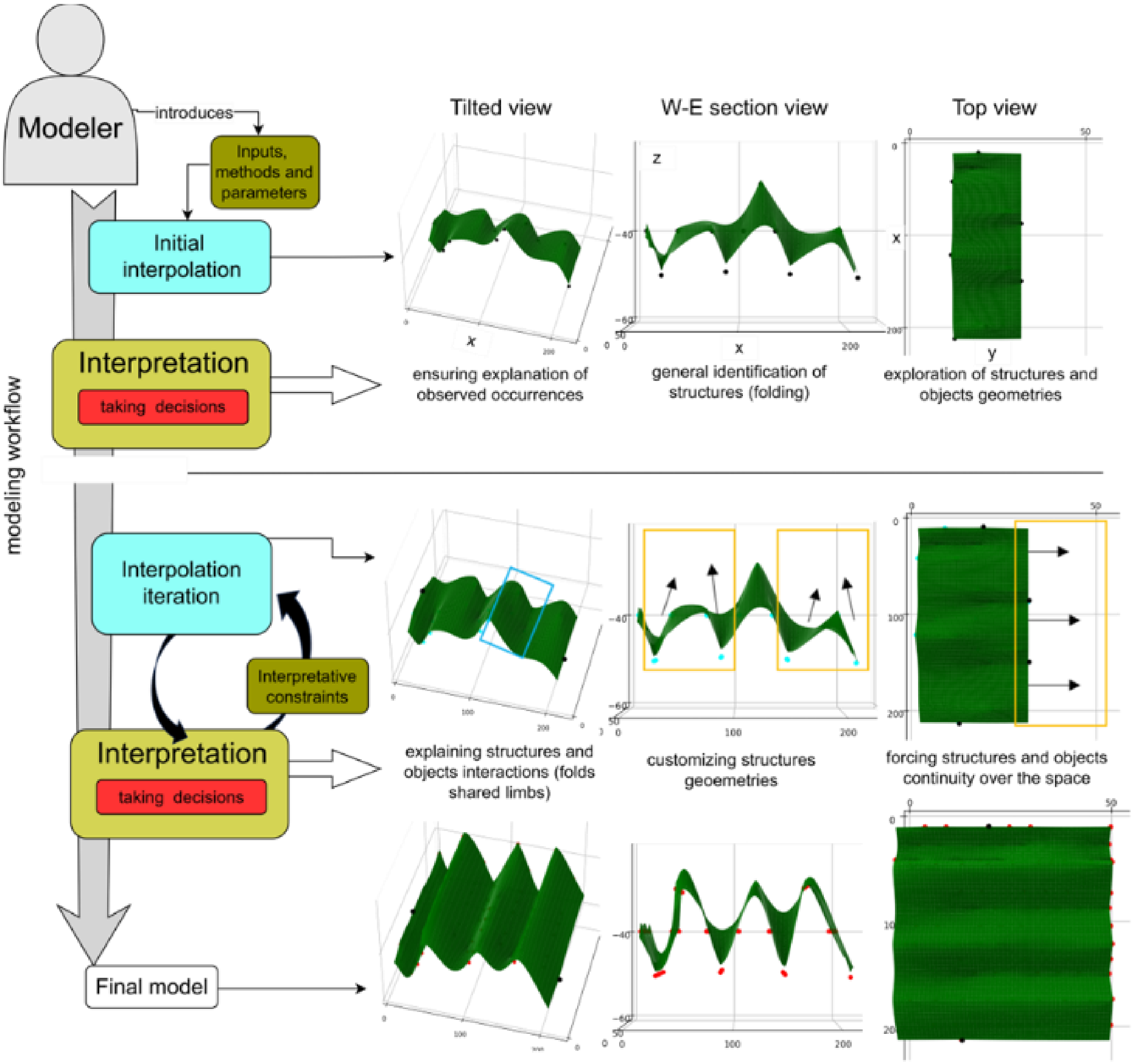


Figure 2-2 Typical modeling workflow undertaken by users to construct surface_1 using RBF for interpolation. Interpolation outcomes are highlighted in green across tilted, section and top views. Black points represent the initial dataset containing 11 points. Cyan points represent 9 control points added in the second iteration, and red points represent the final dataset containing 100 points. White arrows from interpretation boxes indicate key modeler actions. In the second tilted view, a blue inset represents a shared limb between two successive folds. Yellow boxes in the second section and top views denote user-identified zones for improvement in subsequent iterations. Black arrows indicate potential locations for adding interpretative constraints during the process

2.3. A knowledge-driven modeling paradigm: theoretical foundation

As shown by the use case, a critical element in the modeling process is interpretation that leverages knowledge. The next sections present the theory behind the proposed paradigm as inspired from the natural human interpretation process. Two main elements are presented: an interpretation formalism (section 2.3.1) and a knowledge model (2.3.2).

2.3.1 Interpretation formalism

Within the geomodelling context, interpretation is employed to address gaps in information arising from data scarcity and discontinuities in inputs. Examples include isolated stratigraphic parts of the same geological unit, unknown objects or missing properties of (e.g., a stratigraphic part with an unknown stratigraphic attribution), or ambiguous expressions of geological objects (e.g., a tilted sedimentary layer). These challenges are referred to as unexplained situations. They often fail to satisfy key geological principles or basic assumptions, such as the principle of original horizontality, or lack sufficient information required for modeling –such as missing property values– necessary for a geological entity to be considered valid within the modeling framework.

Interpreting these situations involves proposing some entities (objects, relations, or properties) or updating properties of existing entities to provide an explanation within the modeling context. For example, a situation consisting of two isolated stratigraphic parts can be explained through a larger part that covers the gap and initial parts. On the other hand, the tilted sedimentary layer can be explained by the existence of some deformation structure (e.g., fold, fault). This formalism relies on the existence of a finite number of possible explanations for situations. For example, all types of structures that could modify the geometry of a geological unit can be formalized and structured. The operator will create an instance of one of these possibilities and see if it fits the context. Plausibility –impossible to confirm due to the incomplete geological record– is estimated based on respecting data and geological rules.

Interpreting situations involves several stages, which can be broken down as follows. Experts 'plan' the interpretation by selecting an unexplained situation, analyzing it, identifying potential ambiguities, and formulating theories to explain it. Next, they 'do' the interpretation by selecting a theory if multiple options are available. Following this, they 'evaluate' whether the interpretation aligns with their intentions. Finally, they 'act' based on the evaluation: either accepting the interpretation and updating their understanding or rejecting it and moving forward with alternative approaches. These steps can be summarized into four main stages: (1) planning the interpretation, (2) doing it, (3) evaluating the results, and (4) acting based on the evaluation.

Experts repeat this process each time an unexplained situation is encountered continuing until the model is completed. Two notions emerge here: the iterative aspect, to integrate new information; and the incremental aspect, where the process progresses gradually, with new insights. Each iteration adds to the existing understanding, leading to a continuous improvement in the interpretation. When experts conclude an interpretation iteration, they retain information about it to help in further interpretation iterations, for example, to avoid taking a proven dead-end path, which highlights the importance of using information about the interpretation process itself, and not only its outcomes.

The interpretation process also comes into play when experts encounter a situation – resulting from previous interpretation iterations– that does not align with existing information, such as identifying a fold with superposed limbs. In such instances, experts would pinpoint the inconsistency, follow the same process as outlined earlier, and ultimately update or discard the previous interpretation. This aspect underscores the ongoing nature of interpretation, emphasizing its ability to be always updated if new information appears.

Experts also formulate interpretations initially and then endeavor to align them with the constraints of modeling tools. For instance, an expert could mentally think of a fault consisting of multiple segments plus a fault core, give it some properties, and communicate information about it to colleagues, but without representing it mathematically. Then, if needed, the expert would create the geometrical representation that abstracts the fault into one surface. This implies that challenges related to representing a given interpretation should be addressed during the representation phase rather than during the interpretation itself, as a model should ideally remain independent of its representation.

At its conclusion, the interpretation process unfolds as a staged, iterative, and incremental elucidation of a series of situations. Outputs can be modified in light of new information and are represented according to the preferred tool. In our approach, such knowledge retention and updates require support from a knowledge model. This model contains the knowledge used by human or machine interpreters, including pertinent facts, types, and relations.

2.3.2 Knowledge model

The knowledge required by the formalism can be grouped into geological, interpretation, and representation aspects:

- Geological aspects describe the physical geological reality, such as the objects, structures, and their characteristics, e.g., a stratigraphic part is located in space and time; a fold is hosted by a deformed stratigraphic part; a fold has some properties such as wavelength.
- Interpretation aspects include unexplained situations, explaining entities, and a record of each interpretation iteration. Explanation is represented as an explained by relation linking the unexplained entity to an explaining entity. E.g., isolated objects such as initial observations are explained by a larger, complex and semantically consistent object such as some observations are explained by stratigraphic parts, which are explained by a fold if they are deformed. Recording each situation and its interpretation enables history-tracking and tracking of an entity's interpretation status (explained/unexplained) and origin (resulting from interpretation/initial observation).
- Representation aspects provide a spatial and temporal framework to depict geological and interpretation aspects (e.g., representing objects in a given coordinate system, in 2 or 3D; representing only certain aspects interpreted at a given iteration).

One solution for explicitly structuring geoscience knowledge is the use of ontologies (Abel et al., 2012b; Brodaric and Gahegan, 2010; Fernández-López and Gómez-Pérez, 2002; Perrin and Rainaud, 2013). An ontology in information science is a specification of a shared conceptualization (Gruber, 1995) via a logical theory (Guarino and Giaretta, 1995).

It provides a framework to define and organize entities of a domain of interest. An ontology consists of *classes*, that are abstract categories or groupings of entities sharing common properties, such the *Fold* class; *instances*, that are instantiations of classes, e.g. an observed or interpreted fold instantiates the *Fold* class; *properties*, that are characteristics of classes and instances, such as size, shape, or thickness; *relations* that hold between classes or instances, such as a unit being above another unit; and *rules* that are additional constraints on other components, such as older units not being above younger units without intervening factors. While the adoption of ontologies in the geosciences is relatively recent compared to other fields (Dermeval et al., 2016; Ivanović and Budimac, 2014), the results of applying them in geosciences are promising and encourage further progress (Abel et al., 2019; Babaie et al., 2006, 2020; Brodaric, 2012; Garcia et al., 2020; Lombardo et al., 2018; Mantovani et al., 2020).

2.4.From theory to practice: algorithmic implementation

This section proposes a prototype algorithm that implements the knowledge-driven modeling paradigm discussed in Section 3. This algorithm is composed of three main blocks (Section 2.4.1): a Geological Knowledge Manager, an Interpretation Manager, and a Representation and Visualization Manager, as sketched in Figure 2-3. The algorithm relies on the formalization of knowledge aspects using MOGI (Minimal Ontology for Geological Interpretation) (Section 2.4.2). One concern is to ensure the versatility of the developed algorithm and its adaptability to evolving concepts and context. To that aim, several aspects of the algorithm are based on certain distinctions: the separation of conceptual interpretation from the mathematical representation (a) as often done by experts, as well as the separation of tasks from how they are carried out (b).

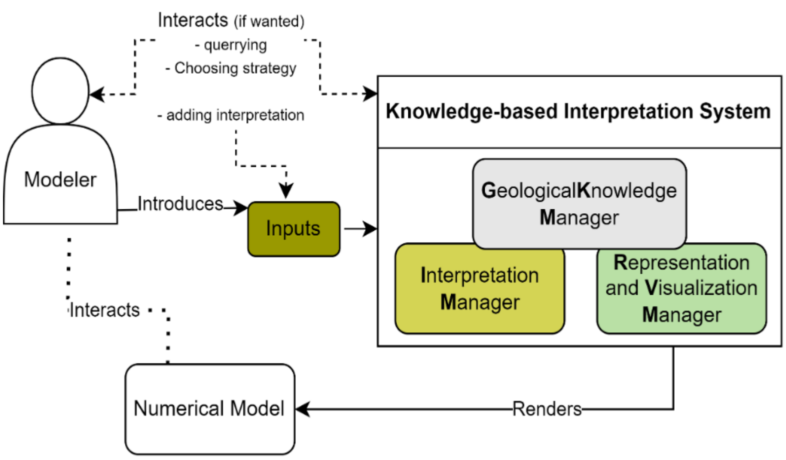


Figure 2-3 Simplified architecture of the proposed Knowledge-driven Interpretation System, which consists of three main blocks: a knowledge manager, an interpretation manager, and a representation manager. The modeler can interact with the system at different levels: by introducing inputs, by adapting the knowledge framework, by giving insights to the interpretation process, and by controlling parameters of the final numerical model

For the former (a), the algorithm relies on manipulating instances of ontology classes that correspond to experts’ mental concepts, and these instances do not need to be represented spatially to exist or to be manipulated. For example, counting the number of fold instances can be done without the need to fully represent each fold in 3D, but when represented the instances can be attached to one or multiple spatial representations. For the latter (b), the different subprocesses run by the algorithm are defined as generic tasks, which can be executed via multiple strategies at run-time depending on user choice, availability, or suitability in certain conditions. For example, selecting some inputs to be interpreted is a task; however, the selection could be done in different ways such as randomly, by type of input, e.g. selecting only field observations of a certain campaign, or by region such as observations located only southern of a latitude line. Users can thus implement their preferred method for performing the generic task or use per default parameters.

2.4.1 Algorithm architecture

The Geological Knowledge Manager (GKM) stores and implements MOGI instances (Figure 2-4). At the initial stage of modeling, it maps data from external databases to new instances of the underlying MOGI ontology. At this stage, only excel sheets and Panda’s data frames are supported. This generic manager defines a contextual knowledge framework that runs during the modeling process.

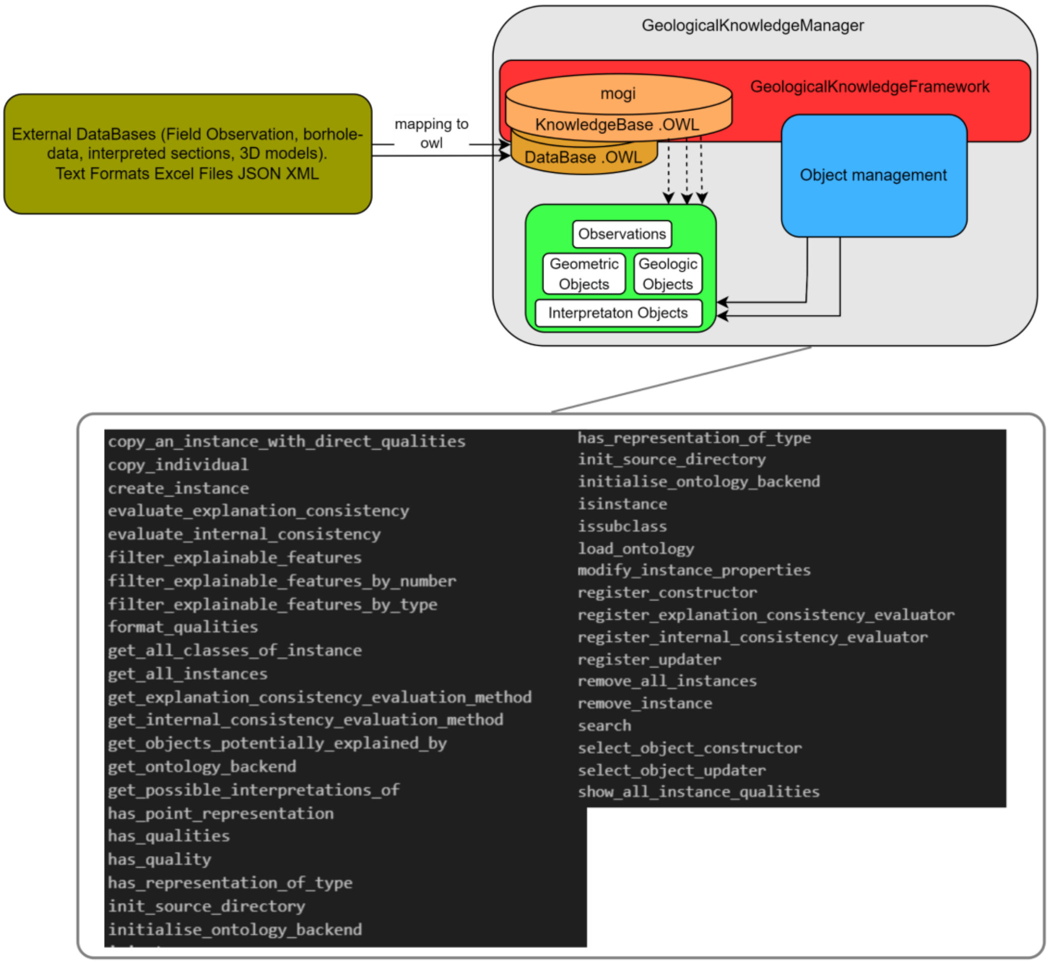


Figure 2-4 Illustration of the main components of the GKM and its interaction with modeling inputs. The lower box depicts the main methods and constructors used to build the GKM

The GKM is designed to be able to include also other ontologies and merge them with MOGI within the same modeling framework. It also provides functionality to build, update, and manage geological feature instances based on the class definitions in the MOGI knowledge model. The GKM's design emphasizes modularity by assigning distinct responsibilities to each method. The operations are abstracted in such a way that new functionalities can be added without altering the core architecture.

The Interpretation Manager (IM) applies the formalized interpretation process on the inputs. This formalized process process is implemented as a continual improvement cycle, for which a Deming wheel formalism is adopted (Deming, 1997). One iteration of the wheel is composed of four successive subprocesses Plan, Do, Check, and Act (PDCA). This Deming wheel is used in quality management and process improvement methodologies, particularly in manufacturing, healthcare, and software development (Moen, 2009). The PDCA cycle incorporates the various interpretation steps outlined in our earlier theoretical analysis. The IM provides a broader framework to manage and executed these processes (Figure 2-5) and employes functions provided by the GKM (e.g., constructors and updaters, Figure 2-4):

- **Plan**: identifies where the process itself is needed by finding and selecting unexplained situations. Situations can be selected following different strategies, e.g., randomly, by localization, or user input. "Plan" also selects the type of action to be applied based on the selected situation, either New, Remove, or Update interpretation, and provides context by listing related information. For example, if a dipping stratigraphic part situation is selected, the related information can consist of a list of candidate explanations (e.g., the existence of a fold). Another situation can be the existence of a stratigraphic part that it is contradicted with an observation (e.g., polarity issue). Hence, if the action is to update it, related information will include querying the range of possible values of the polarity of the stratigraphic part.

- **Do:** applies the planned action. If a new interpretation is to be created, "Do" chooses the class of object, either randomly or via another strategy, and creates a new instance of this class and estimates property values. This instance creation and value estimation is referred to as realization. Realization can be derived from quantitative analyses of available features (e.g., inferring a dip from a series of contact observations) or from general geological principles when available observations are not informative (e.g., quantifying a fold limb parameters based on a the parameters of the first limb already interpreted). Applying Remove or Update will seek methods from the Geological Knowledge Manager to safely remove the selected object or update its properties. The proposed interpretation remains hypothetical and will be accepted or rejected in "Act" after its evaluation in "Check".

- **Check:** evaluates the proposed or updated interpretation for compliance with stored knowledge. In strict compliance, the interpretation satisfies all evaluation criteria. Three classes of criteria are used: internal consistency refers to compliance with valid internal properties (e.g., a fold should have exactly two limbs, not more or less); external consistency refers to compliance with valid relations to exterior objects (a new identified stratigraphic part should not cut another part); and explanation consistency means the proposed interpretation must explain its situation. Failing to respect one of these criteria generates an anomaly. The algorithm is designed to be flexible and accommodate different strategies for evaluation, such as partial compliance with only some criteria, or using some probabilistic method. If the check fails, then the existence of an anomaly is flagged, but an instance of Anomaly is not yet created (it is created during the Act step). An anomaly does not necessarily signify the interpretation is wrong, as the anomaly itself becomes a situation that needs to be interpreted in further iterations.

- **Act**: finalizes an interpretation iteration based on the checked results. The interpretation will be validated or removed, anomalies will be created if they are flagged, and if necessary, the need for a new iteration will be designated. To show how the existing understanding is evolving and updated, the algorithm also computes the spatial coverage of the interpretation over the studied area. The user also can interact with the system at this stage, by stopping the process or querying for the status of the interpretation or status of a given situation.

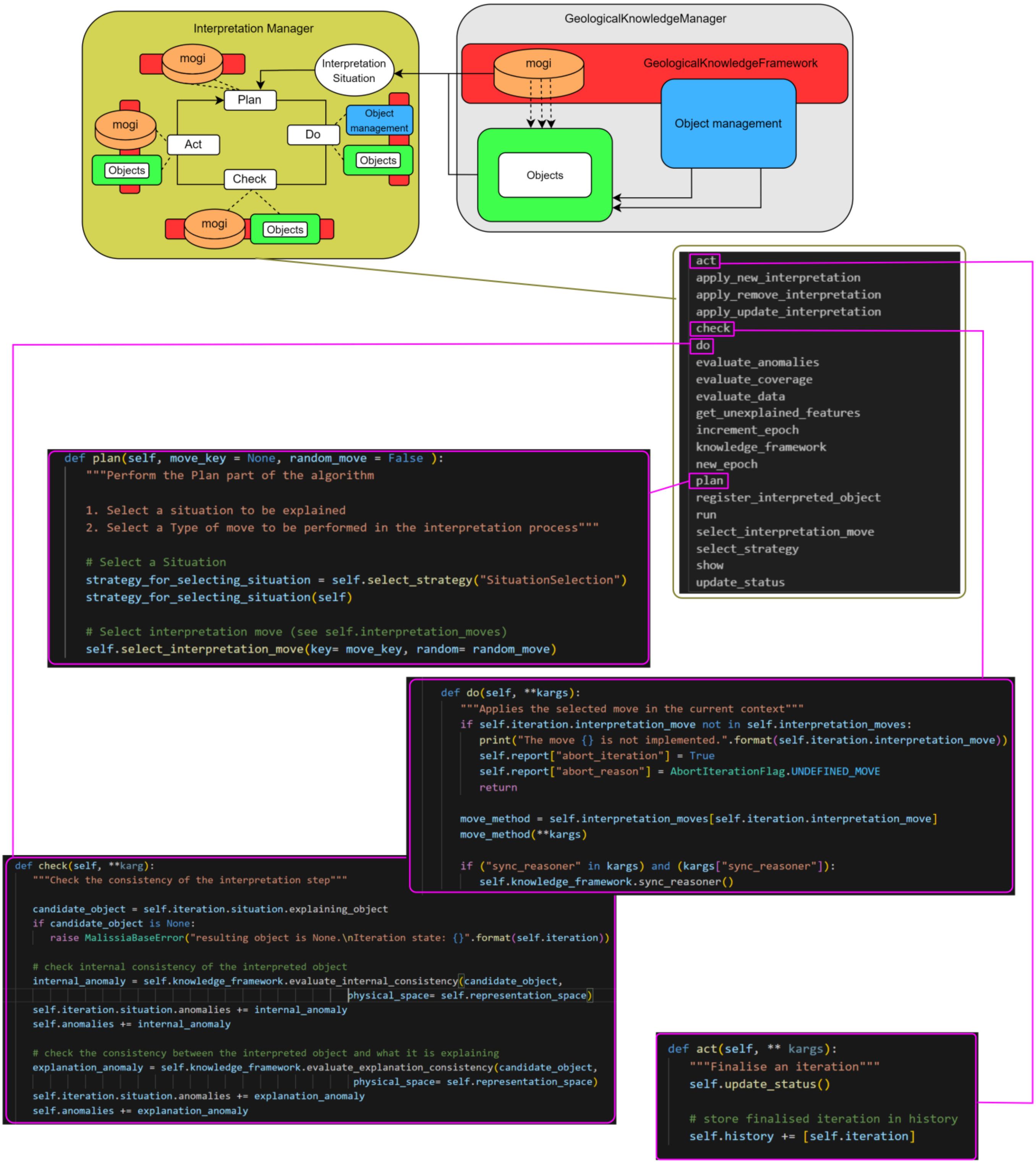


Figure 2-5 Illustration of the main components of the IM and its interaction with the GKM. The lower boxes depict the main methods and constructors used to build the IM

The Representation and Visualization Manager (RVM) enables the creation of various spatial representation spaces for interpretations (Figure 2-6). These representation spaces allow different interpretations or models to be visualized and analyzed. A model can be associated with multiple representation spaces, each having its own unique set of parameters, such as coordinate systems, resolution, scale, and 2D or 3D.

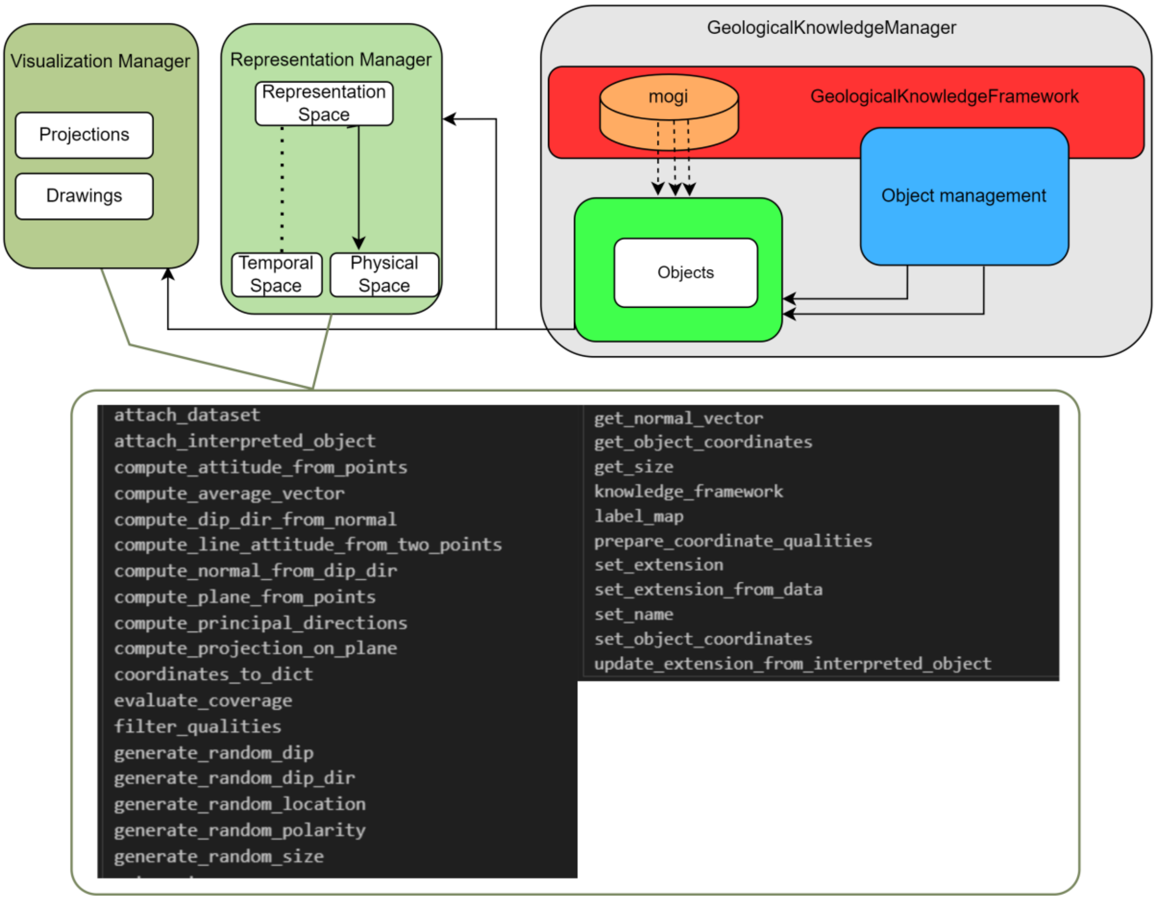


Figure 2-6 Illustration of the main components of the RVM and its interaction with the GKM. The lower box depicts the main methods and constructors used to build the RVM

2.4.2 Knowledge model: MOGI

The different knowledge aspects required for interpretation (Section 2.3.2) are encapsulated in an ontology referred to as the Minimal Ontology for Geological Interpretation (MOGI). "Minimal" indicates that ontological characterization of these concepts is kept to a minimum to focus on demonstrating the feasibility of the interpretation algorithm. MOGI concentrate on structuring key-knowledge elements about the description of folding patterns, particularly chevron folding (Ramsay, 1974). MOGI is fully described in appendix A (with github link) and only its main features are reported in this section (Figure 2-7). Four principal entities are defined in MOGI:

- Geological_Object: includes classes for basic geological entities: Stratigraphic_Part, Stratigraphic_Surface, Fold, Fold_Limb, Fold_Train, Fold_Axis and Structural_Parameter.
- Representation_Object: includes classes for representing each class of Geologic_Object.
- Observation_Object: includes one class: Pointbased_Observation.
- Interpretation_Object: includes the Situation and Anomaly classes. Two anomaly subclasses used in the examples are Dipping_Anomaly and Dip_Variation_Anomaly, and both are related to stratigraphy. Dipping_Anomly refers to the situation of having one stratigraphic part tilted, on the other hand, dipping_Variation refers to the situation of having a stratigraphic part with some variation of the dip over its area. Explanation_Anomaly is also used in the examples and is generated when an interpretation is proposed but does not explain all the features in the selected situation.

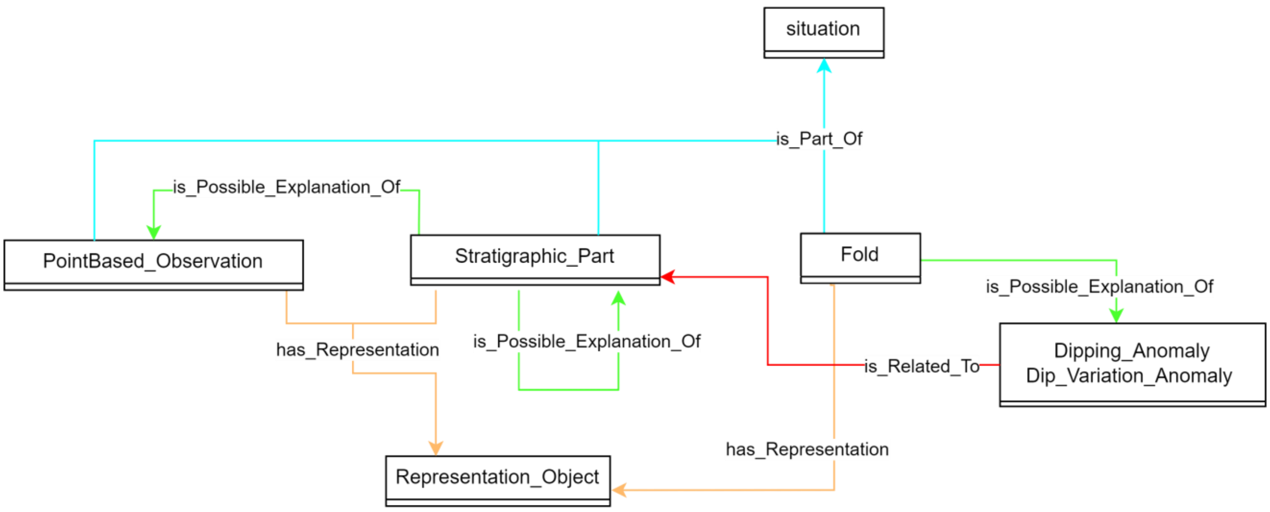


Figure 2-7 Main classes and relations in MOGI

.

2.5. Results of application

The implemented algorithm is applied on two use cases: on the synthetic data used earlier (section 2.2), and on a set of real data from the Center Armorican Domain (CAD), in Britanny region, western France.

2.5.1 Application on synthetic data

The figures 2-8 to 2-11 highlight key situations central to algorithm functionality. Each demonstrates an interpretation iteration automatically executed by the algorithm without expert involvement. These situations are projected onto a 2D vertical cross section for illustration purposes, but the interpretation considers their three-dimensional properties. A final 3D visualization is presented in Figure 2-12 to compare the results of the algorithm with separate interpolation results. To avoid redundancy, the PDCA cycle is explained in more detail for the first and second iterations, with explanation of subsequent iterations focused on the interpretation aspects.

Iteration 1

- **Plan**: Figure 2-8a presents inputs (in black), with a portion (in blue) selected for an interpretation iteration using a random selection process. Based on the type of input (observations), the algorithm selects the only existing possible explanation defined in the knowledge base that can explain this type of situation (a stratigraphic part).
- **Do**: Figure 2-8b shows a realization of the stratigraphic part class (in green and yellow), derived from an estimation process based on the average values of spatial properties of the selected inputs.
- **Check**: Figure 2-8c shows the proposed stratigraphic part does not pass Check (in green and red), because it does not explain all the candidate inputs (the unexplained input is also in red) based on the validation criteria (spatial properties such as the localization of the proposed part compared to existing parts, or its polarity).
- **Act**: Figure 2-8c also shows an anomaly is flagged (in red) (Explanation_Anomaly). This anomaly represents a new situation for further interpretation - it could be modified to fit observations or deleted if there is no possible good fit.

Iteration 2

- **Plan**: Figure 2-8d show different inputs (in blue) selected for interpretation. Like the first situation, the only existing possible explanation from the knowledge model is a stratigraphic part.
- **Do**: Figure 2-8e shows an instance of the stratigraphic part class, built using the average estimation process.
- **Check**: Figure 2-8f shows the stratigraphic part passing Check as it respects the evaluation criteria.
- **Act**: in Figure 2-8f the interpretation is validated, and the interpretation status is updated by computing the area covered by interpreted objects (green dashed line).

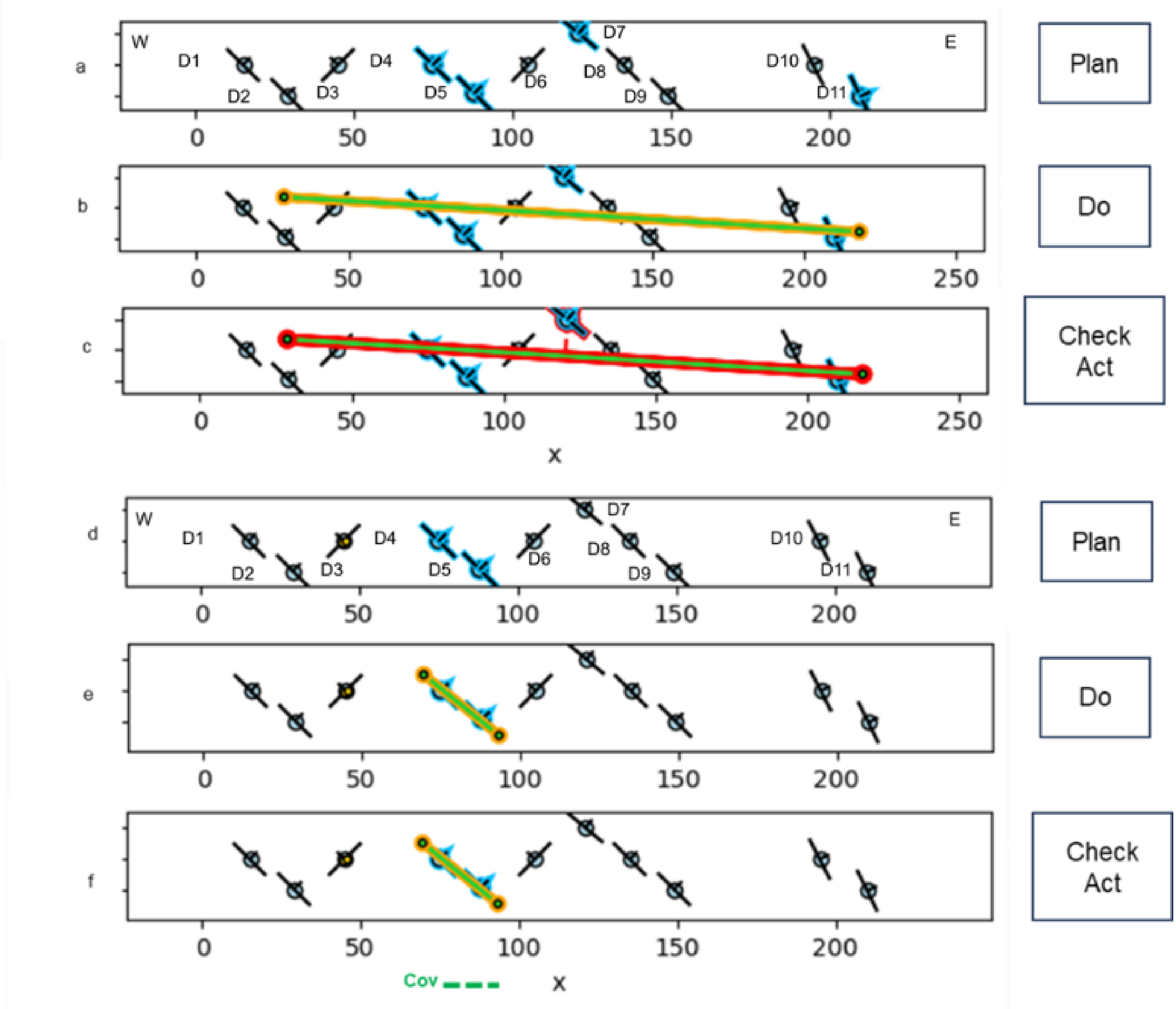


Figure 2-8 Two full iterations of interpretation based on the PDCA. Features selected in each iteration are in cyan blue. Proposed explanation objects are in green. Failed checks are in red. Dashed green line indicates a coverage update. See text for details

Figure 2-9 shows a situation in which two stratigraphic parts – one input (D3) and one interpreted in previous iterations (line in blue)– are selected (Figure 2-9a), explained (in green and yellow, Figure 2-9b), and checked. A Dip_Variation_Anomaly is flagged during checking, because the stratigraphic parts explain all objects in the selected situation, but the parts have different dips (in green and red; Figure 2-9c). In further iterations, this anomaly is selected and interpreted as a fold with an axial surface (blue line). The resulting fold object passes checking, an update of the system saves the fold as a validated object (in yellow, Figure 2-9d) and the green line indicates increased spatial coverage of the studied area by the interpretations.

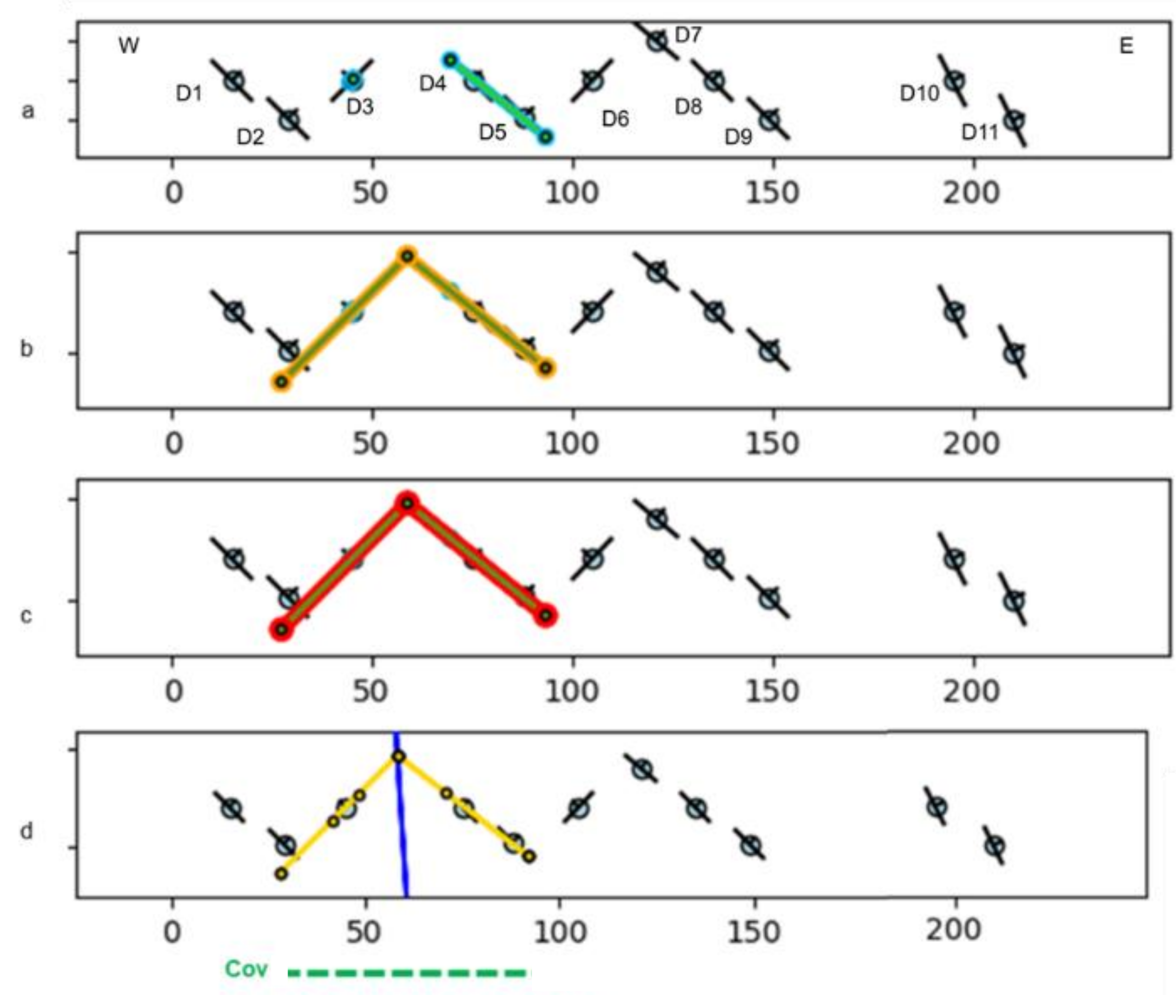


Figure 2-10 interpretation of a deformed object. See text for details

Figure 2-10a depicts the selection (in green and blue) of two stratigraphic parts Strat_P and Strat_F (Strat_F is the fold interpreted in Figure 2-9), and Figure 10b shows their interpretation (in green and yellow) as a deformed stratigraphic_part with adapted geometry and topological connections. Dashed green line with "cov" symbol indicates the updated spatial coverage of the studied area by the interpretated objects. Blue lines are finite representations of fold axial surfaces.

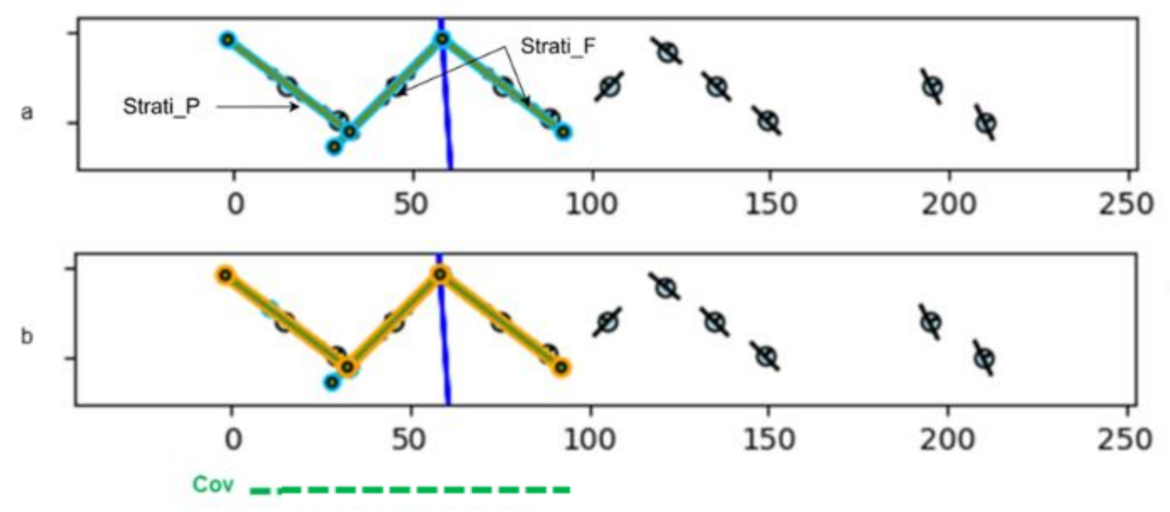


Figure 2- 10 interpretation of two stratigraphic parts, one with simple geometry, Strat_P, and the other with complex geometry, Strat_F. See text for details

Figure 2-11 shows the final iteration, in which the algorithm explains all situations and the studied area is entirely covered by the interpretations (green line). Stratigraphic parts (in yellow) are created to explain observation data, or to explain other smaller stratigraphic parts created during the interpretation process. Folds explain deformed stratigraphic parts (each fold with its axial surface in blue) (to avoid distraction in the 2D section only axial surfaces of each fold are shown, see Figure 2-12 for the full 3D overview), and a fold train object (the entire structure in yellow) explains the existence of a series of successive folds.

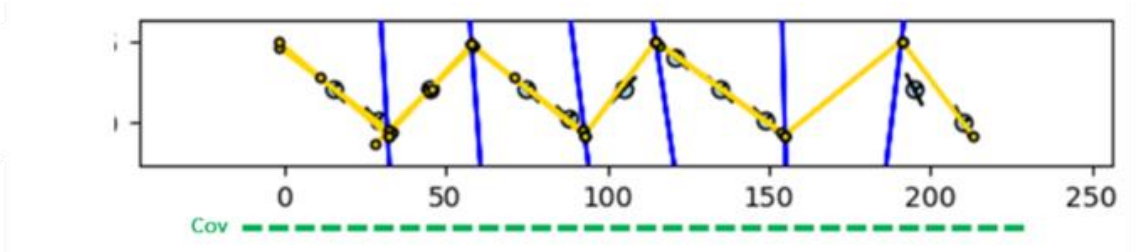


Figure 2-11 Final results with explaining / interpreted objects highlighted in yellow. See text for details

Figure 2-12 compares the model resulting from RBF interpolation (Figure 2-12a, the same model resulting from interpolation in the use case Figure 2-1) with our results (Figure 2-12b). Both models are represented in 3D and exhibit similar overall shapes, with notable distinctions. The interpolated model lacks associated descriptive information such as various properties (axial surface, fold axis) of the created fold instances. Moreover, in areas without data (e.g., the left and right sides of Figure 2-12a), the interpolated model appears smoothed and flattened, particularly at the hinge. This smoothing is consistent with a continuous geological surface, but potentially overlooks other relevant knowledge not included in the interpolation, such as the fold objects. In contrast, this and other knowledge, such as theories, principles and rules, is stored by the GKM and used to guide the interpretation: e.g. properties and parts of the folds (axial surfaces, axes of folds, limbs) are stored in the knowledge base and could be queried throughout to guide the process (Figure 2-12c). This enhances control over outcomes in regions of high epistemic uncertainty.

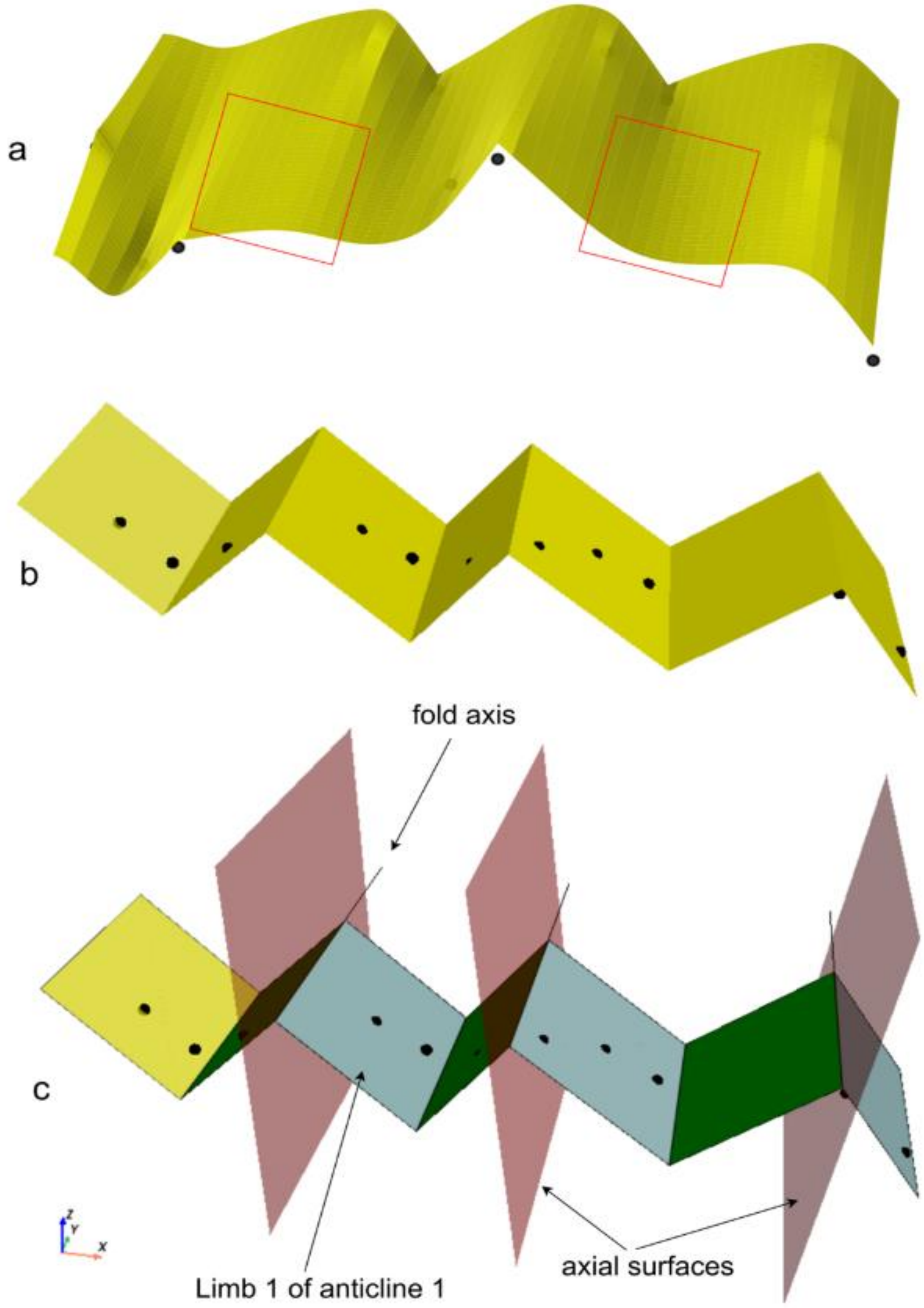


Figure 2-12 Comparison of standard RBF interpolation (a) and our results (b) on synthetic data. The red frames in (a) highlight oversmoothed areas in the interpolated model. Various properties of the interpreted are shown in (c). These properties are attached to the model and can be queried by experts: anticlines are shown with their axial surfaces, as well as axes and fold limbs. Each limb is colored based on its vergence with respect to anticlines, with the yellow part disassociated from an anticline.

2.5.2 Application on the CAD

In the second use case, the algorithm is applied to a more complex and realistic dataset from the Central Armorican Domain (CAD), whose geology was previously discussed in Chapter 1, section 1.1.3. This application focuses on delineating the boundaries of key geological units: the base and top of the *F. de Grès Armoricain*, the top of the *Traveusot Chatelier Fm*., and the top of the *Riadan Formation* that belong to the Ordovician group. These units present an overall folding with low to medium amplitudes (dipping between 20° and 40°). To focus more on how the system handles interpretations, the dataset was manually extracted from the geological map (Figure 2-13) with some simplifications, such as ignoring fault offsets and enforcing a consistent deformation pattern across observations to minimize heterogeneity. Similar to the first use case, each interpretation iteration, following the PDCA cycle, was executed automatically by the system.

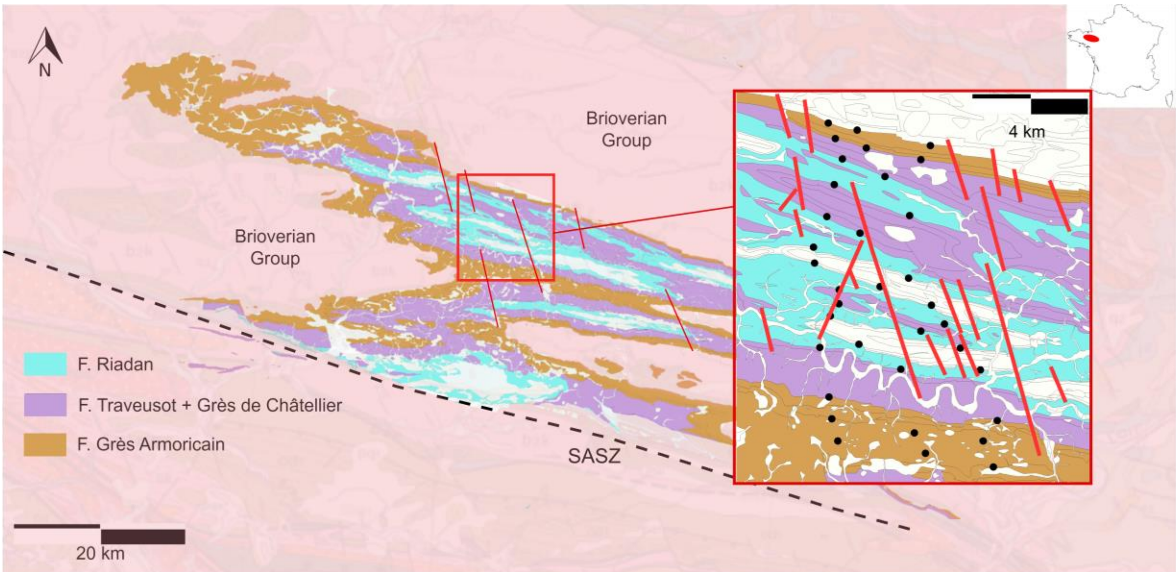


Figure 2-13 Geological map of the Brittany region where observations are taken. The red frame highlights a zoomed-in area where observations (black dots) for the application are sampled. Samples are primarily taken along the boundaries of key geological units targeted in the study (the *Grès Armoricain*, *Grès de Traveuséot*, *Grès de Chatellier*, and *Riadan* formations) original map at BRGM and further details in Figure 1-5

Figures 2-14 and 2-15 illustrate various snapshots captured during the model construction process across multiple interpretation and representation iterations. The adopted workflow in this application follows a chronological order for constructing the structural surfaces. Figure 2-14 presents the result of the modeling of the base of the *F. de Grès Armoricain* after several iterations of interpretation, starting by multiple local stratigraphic parts (Figure 2-14A and B) and progressing to a more regional interpretation that contextualize and explain these local interpretations (Figure 2-14D and E). The final generated interpretation is depicted in Figure 2-14I and G. These steps of interpretations are summarized in an interpretation history log (Figure 2-14C and F) highlighting key features, the context of the interpretation, possible candidate explanations, the resulting interpreted object, and its evaluation –flagging any detected anomalies.

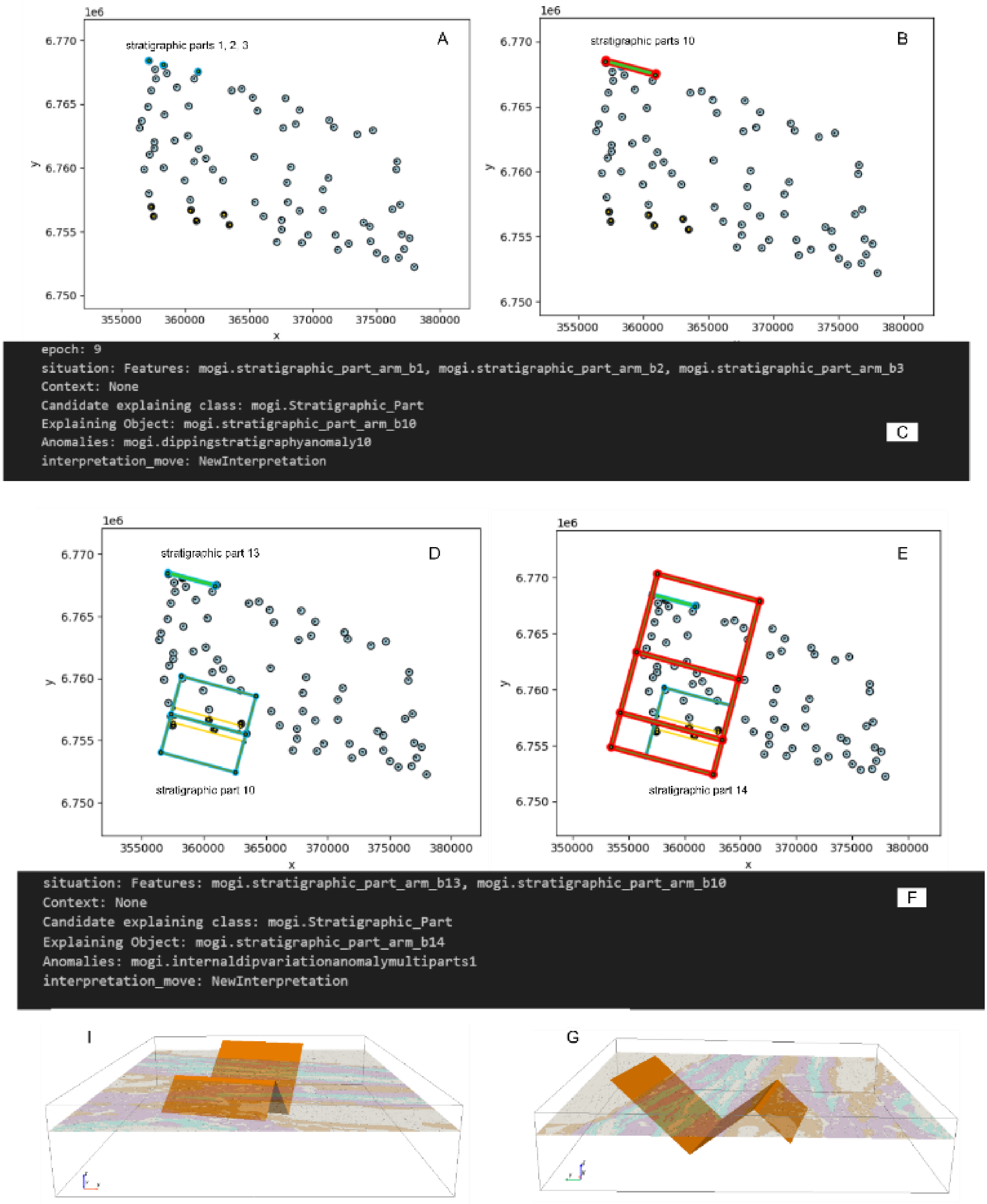


Figure 2-14 Illustration of the modeling of the Base of the F. Gres Armoricain unit through various iterations A to F. inputs (in Cyan blue) are selected in A and interpreted in B (highlighted in green and red). C shows the history of this interpretation iteration. D shows a subsequent iteration where interpretation from previous iterations is selected (Cyan Blue and interpreted in E (green and red) with the history in F. Final constructed model of this unit is displayed in 3D in I and G with different visualization angles

Figure 2-15A displays the final interpretations of the structural surfaces influenced by various folding structures. Figure 2-15B demonstrates the update of these folds, adjusted to fit topographic constraints via the application of an erosion surface. Although the resulting model is a rough approximation, including the geometries of units, spatial distribution, and structural properties, these results underscore the significant contribution of the developed algorithm in addressing the primary scientific objectives and requirements. In fact, it allows to have a framework where the interpretation process (suggestion of structures, modifying their properties, omitted anomalous entities) is completely transparent. This framework is interactive as users can interact with the different interpretation steps by defining how the system can perform tasks, performing them if they want, or simply query the system in each step.

Essentially, the interpretation and the modeling processes, through the proposed system, are: (1) semantically rich and logically consistent, (2) capable of leveraging geological classifications, theories, principles, and rules defined in an external knowledge model to build, update and validate the constructed models, and (3) able to provide clear explanations for decisions made throughout the model-building process.

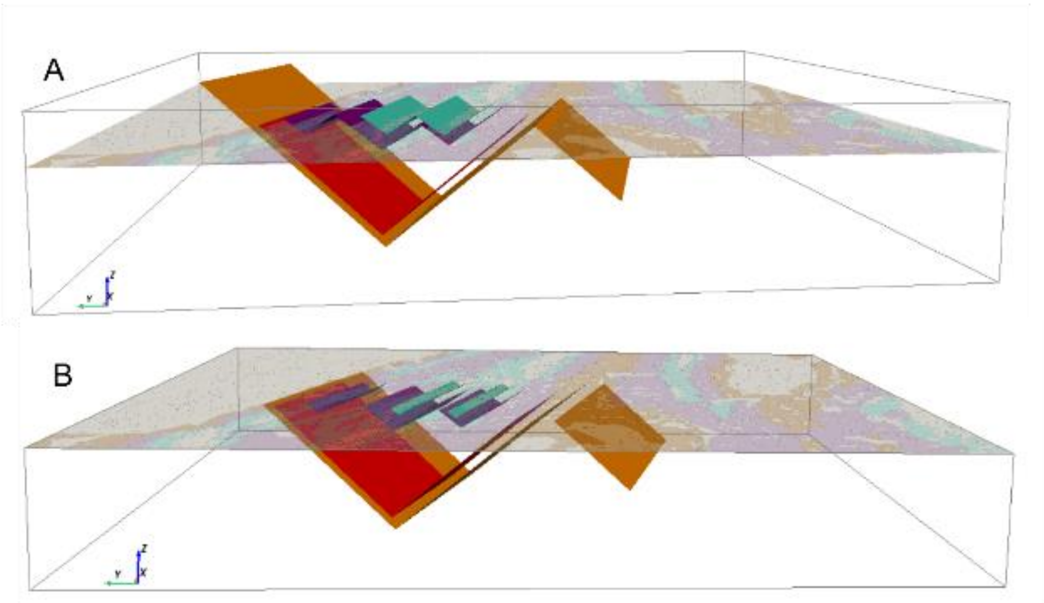


Figure 2-15 Final 3D model of the study area with the four structural interfaces. A displays the original geometries resulting from interpretation, and B displays these surfaces after trimming via the application of the erosion surface (topography)

2.6. Discussion

This section revisits the key challenges identified in the introduction: automatic interpretation via the use of knowledge, the preservation of the knowledge in the model and its representations, and associated uncertainty. Although these aspects are addressed variously in the literature, they are not addressed holistically as per our approach.

2.6.1 Achievements

2.6.1.a A formal interpretation paradigm: important steps toward automation

The formalization and automation of the interpretation process is an innovation in the field of geological modeling. This automation covers interpretation processes usually carried out by experts: from input selection, theory proposition, to quality control. These processes are applied during model building for the identification and classification of objects, and construction of geometries and relations. A particular advantage is the generation of hypothesis, which is not typically part of classical interpolators. For example, folding deformation is automatically proposed in the case study without the need for its identification by experts (Grose et al. 2017, 2018).

Although the approach aims to automate the interpretation process, it does not totally exclude the intervention of experts. Experts could for example intervene to perform a certain task (e.g., forcing the selection of a situation to be interpreted). Experts also could intervene by injecting their interpretations in different formats, such as ontology entities or 3D objects. The proposed prototype is able to convert them, extract semantics, and integrate them in the knowledge base. As these interpretations are introduced with their full semantic content, there is no loss of knowledge, in contrast to classical interpolation-based modeling in which most common interpretations are used merely as interpretative constraints to force geometric representation (Lemon and Jones, 2003).

2.6.1.b Representation of the model and knowledge retention

Although the case study focuses solely on folding and unit boundaries, it separates the model from its geometric 3D representation, with the geological properties attached to the instances of objects in the knowledge model and the representational properties attached to the geometric representation. This enables semantic manipulation of the objects, and allows for downstream preservation of the knowledge, beyond the mathematical equations defining the structural surfaces. This coincides with interest by the research community in storing knowledge for exchanging 3D models (Legg et al., 2015; Morandini et al., 2011; Z. Wang et al., 2018; Zhan et al., 2022). In addition to fully preserving information about geological and representation aspects, the formalization of the interpretation process enables the recording of an interpretation history. This not only includes tracking the interpreted elements, but also the reason behind each choice.

2.6.1.c Uncertainty in the model

According to Wellmann and Caumon (2018), uncertainty is present at all stages of 3D model development and it originates from different sources. Extensive effort has gone into developing methods for quantifying uncertainty, particularly for the measurement of geological objects, and the choice of mathematical framework and interpolation methods (e.g., Chilès et al., 2007; He et al., 2015; Jessell et al., 2018; Jørgensen et al., 2015; Liang et al., 2021, p. 20; Olierook et al., 2021; Pakyuz-Charrier et al., 2018; Schweizer et al., 2017; Wellmann et al., 2018). Uncertainty related to the understanding and conceptualization of experts –known as epistemic uncertainty– is also an active area of study (Bond, 2015; Bond et al., 2007; Heim et al., 2019; Liang et al., 2021; Ulmer et al., 2021). However, capturing and quantifying epistemic uncertainty remains challenging, as it is related to the subjective understanding of experts, including how they explicitly and implicitly use knowledge (Wellmann and Caumon, 2018).

Although the present paper does not specifically focus on uncertainty, our differentiation between a model and its representation begins to separate semantic concerns from spatial representational concerns, including related uncertainties: e.g. the classification of an object as fold might be valid, even if its geometric representation as a line or surface is suspect. Moreover, the self-anchoring (Bond, 2015) –where experts become attached to their initial interpretations, limiting their ability to consider alternatives– is indirectly addressed by the system. When experts receive suggestions from the system rather than producing them, they are less likely to become attached to a specific idea, encouraging them to take a step back and look to the bigger picture to evaluate multiple possibilities.

2.6.2 Prospects and areas for improvement

This first part of the research work paves the way for the formalization of expert's knowledge used in the 3D modeling process by identifying main knowledge aspects (geological, interpretation, and representation aspects). While the focus was on demonstrating the feasibility of the proposed approach, these aspects were structured in a simplified manner within MOGI, without delving deeply into the underlying ontological considerations of these concepts.

2.6.2.a Representation vs. geometric description

A key aspect that requires further exploration is the distinction between a model and its geometric digital representation. In MOGI, the model is defined as a combination of geological entities and their representational concepts (e.g., points, surfaces). Spatial properties, such as dip or size, are associated only with the representational objects. Clearly, such approach is for simplification purposes, though it represents a first step in separating the conceptual understanding of experts from its potential representation. However, more attention is needed to the representational properties of models, as geological objects are physical entities inherently defined by certain properties and values, not just their representations. These inherent properties should be distinguished from those used solely for representation, as the former remain constant, while the latter may vary. For example, a geological model can be represented in various ways, including non-figurative forms like textual descriptions.

Moreover, two 3D models may describe the same geological concepts but with notable differences in terms of properties of this concept –are these the same models, or distinct ones? How do they relate to the underlying conceptual model? Also, when experts propose a simplified version of reality (as often the case) how to explicitly state the criteria used to create this simplification? Or when an expert builds a simplified model of the same study area based on a previous model, how should this relate to the original model? While these questions may seem intuitive to experts, they carry significant weight in computational frameworks, particularly in terms of information storage, knowledge base definition, and how models should be queried or updated when new information emerges.

2.6.2.b Interpretation and its relation to geological entities

Another important aspect to consider is knowledge related to interpretation. MOGI provides two possibilities to deal with interpretation aspects: the explanation relations between explainable and explaining objects (*explains, is_Possible Explanation*), and the anomalies that define situations requiring interpretation or further handling. Besides the explanation notion that itself needs to be thoroughly revised to determine its relation to geological entities, there is an ontological challenge here to relate the entities that stand for the real objects, the model and the different realizations and representations of the model. If certain characteristics change or are found to be incorrect, does the object remain the same, or does it transform? Since the object is supposed to represent a real-world entity, it should not be erroneous, but the model might be. This necessitates a deeper exploration of the inherent relations between the model, its different representations, and the real-world geological entities it aims to describe.

2.6.2.c Processes

A third critical aspect is that the knowledge model MOGI did not focus on referencing the entire processes and interpretive tasks as instances in the ontology, although they are thoroughly documented in the modeling system. One of the challenges is the organization of these entities in the knowledge model, with consideration of what are the major processes, their subprocesses, how they are related to each other, which entities they interact with (e.g., inputs, outputs), and how to ontologically describe them in a generic way to allow for the interpretation algorithm to run them differently.

2.6.2.d Further enhancement

In addition to these critical areas of development based on further conceptual modeling and standardization, there is a clear need for further enhancement of the knowledge model to cover a broader range of geological entities, enabling more realistic applications. Importantly, these geological entities need to be defined based on their geological essence, independent of their possible interpretations or representations. In addition, the prototype algorithm also requires further development, particularly in addressing key aspects of the interpretation process. This includes improving the retroactive capabilities of the algorithm –allowing it to backtrack when the interpretation proves incorrect– and better capturing the strategies experts use during interpretative tasks. The representation aspects will also require further attention to better handle geometric and topological properties of modeled geological entities.

I believe that strengthening these aspects will depend on a more robust and consistent knowledge framework, structured around the considerations discussed above (processes, representation vs reality, and interpretation and its relation to geological entities).

Decisions trees

A first important aspect that requires further development is the retroactivity of the interpretation process. Specifically, when the algorithm reaches a dead-end or encounters a non-plausible situation downstream in the modeling process, there is a need to determine to what extent the algorithm can backtrack to omit an incorrect interpretation, and all subsequent interpretations based on it. Currently, the proposed algorithm primarily focuses on forward interpretations–which means omitting a non-valid interpretation if instructed so. However, it does not automatically delete previous incorrect interpretations. From our point of view, this challenge can be approached by adopting and adapting the decision trees perspective coupled with a checkpointing technique.

Decision trees (De Ville, 2013), often used in mapping out different possible scenarios and their consequences in domain such as finance, healthcare, and machine learning, are hierarchical treelike structures to organize operations (nodes) representing decisions, and their possible consequences (branches) connected to previous nodes. Checkpointing is a technique that consists of saving some ‘states’ when executing an algorithm, enabling it to restart from these points in case of reaching an impasse on failure instead of starting over (Toueg and Babaoglu, 1984). The idea here is that when using the proposed formalism and the prototype algorithm, a decision node could be defined as the situation requiring interpretation and the paths are the different possible theories possible to explain this situation. To define checkpoints, it seems convenient to use decision nodes as points to restart from and explore different paths guided by new conceptual theories. However, it seems more convenient to define checkpoints only at situations that could be explained by multiple theories. This means that if a situation that has only one possible theory to explain it, and the theory was tested, the algorithm must go backward to the interpretation (n-1) that generated this situation and explore from there.

Checkpoints could also be set at the ‘plausible’ model state before the selection of the interpreted situation that led to a dead-end, allowing exploration of other possibilities of selecting other situations and approaching the issue from a different angle. I believe however, Nodes need to be defined at the conceptual level, when proposing instance and before realizing them and concretizing their space-time properties. This way avoids including all properties of geological objects and structures, which would exponentially increase the number of possibilities and transform the model into a highly sensitive and unstable regression tree. Such instability could lead to an endless number of nodes for the same conceptual model due to frequent small adjustments in 3D structural modeling. Pruning and memorization techniques could further enhance the management of the discussed retroactive aspect. However, I believe that the definition of these checkpoints should be handled based on the fundamental idea of separation and genericity (like how interpretive tasks were handled), with a predefined common objective specification that allows flexibility in how checkpoints are established.

This flexibility becomes more valuable when the proposed prototype algorithm is updated to account for a broader range of geological situations requiring interpretation, including all relevant objects and structures. As the number of situations increases, so do the complexity and the number of possible alternatives for defining checkpoints, leading to an NP-hard problem. Another angle to tackle this retroactive aspect is making more intelligent and geologically informed interpretation decisions leading to produce valid models, thus, more robust checkpoints, and reducing the need for retroactivity. This aspect also depends on the modeling objective –whether it is to find an acceptable model for a rapid solution or to explore various possibilities and epistemic uncertainties, considering how interpretation decisions interact with one another, akin to principles in game theory (Myerson, 2013). Ultimately, defining intelligent decisions or specific actions following the intended objective, is done through the definition of strategies.

Strategies

Another area of development to be explored is the strategic aspect that experts employ. For example, the current algorithm proposes to select inputs randomly, though it provides a framework to define other ways of applying this process by users. This process and in general all geomodelling processes could benefit from enhancing the strategic aspects of applying them according to existing information and the context of modeling. For instance, by defining simulation parameters and giving more weight to one way of performing a process compared to another and considering that this weight could evolve as the run. Such strategic aspect extends to more fundamental processes that are used in interpreting a situation: adding, modifying or deleting entities.

Beyond deciding which of the fundamental actions to apply to a particular situation, experts often follow an implicit order of priority for addressing situations, especially those that arise from previous interpretations. These decisions are often based on a multitude of factors and modeling conditions, such as the availability of information, the progress in the modeling process, and, most importantly, the types of situations based on their alignment with geological norms. Many of modeling tools abord this challenge, implicitly from the angle of modeling workflows. One idea proposed in early stages of this research was to conduct a large-scale survey of 3D modelers –including juniors, experienced, academics, and industry professionals – to derive expert-driven strategies for handling situations and priorities for the interpretations process. However, this experiment was not pursued due to the lack of a robust infrastructure focused on the interpretation process and interpretation situations, and the use of explicit knowledge, which have now become more mature.

2.7. Conclusion

This portion of the work introduces a novel approach to geomodelling by formalizing the interpretation process, addressing longstanding challenges such as automating interpretation and minimizing the need for extensive manual expert involvement in model construction, validation, and representation. Experts can engage with the geomodelling process to contribute insights on inputs, knowledge, and interpretations, with the system retaining this knowledge and associating it directly with the model.

The interpretation process is described in an explicit and generic way, and is implemented as a continuous improvement Deming wheel: Plan, Do, Check, Act. The algorithm is implemented in three main blocks: the Geological Knowledge Manager, the Interpretation Manager, and the Representation and Visualization Manager. An important component is the Minimal Ontology for Geological Interpretation (MOGI), which is a very simple ontology for representing the minimal suite of knowledge required and acquired by the interpretation process. Further development of this knowledge model must address the knowledge aspects evoked in geological modeling from a semantic and ontological side while considering their geological properties and representativeness properties. In addition, the origin of knowledge aspects during the modeling must be considered as some of these aspects are acquired through interpretation, thus, exposed to modification, while other aspects are only used as assertions. Subsequently, the formalization of these knowledge aspects cannot be dissociated from their epistemological side.

Chapter 3: Knowledge formalization

Contents

Abstract

Building on the foundation laid in the previous chapter, this chapter introduces an ontological framework to formalize key knowledge aspects employed in the 3D structural modeling, with the focus on those often used tacitly by experts. The objective is to formalize how experts interpret models and their representations and to provide a structured framework for describing the associated processes, which are often implicit. This formalization results in a more explicit and systematic representation of knowledge, contributing to the automation of the interpretation process. The proposed implementation, POKIMON, is a consistent ontology that builds on existing top- and mid-level ontologies, adapting them to specific uses. It is designed to integrate with domain-specific ontologies that describe geological phenomena, while POKIMON itself is designed to formalize the knowledge used to build models representing these phenomena.

This chapter is a preliminary version of a research article (POKIMON: an ontology for geological interpretation in 3D geo-modelling) to be submitted to one of these journals: Computers & Geosciences, Geoscientific Model Development, or Applied Ontology.

Authors and co-authors of this article are:

Imadeddine LAOUICI, Boyan BRODARIC, Christelle LOISELET, and Gautier LAURENT

3.1. Introduction

3D structural geological models (Mallet et al., 1989) are three-dimensional digital representations of subsurface geological architecture. They are the result of a geomodelling workflow, which involves a series of operations (Perrin et al., 2005; Perrin and Rainaud, 2013; Wellmann and Caumon, 2018) typically carried out by multiple geoscientific experts. These experts iteratively interpret inputs (Bond et al., 2007; Zhang, 2008) and configure algorithms to generate numerical 3D representations (Laurent, 2013; Terrington et al., 2008). Through this process, modelers make critical interpretive decisions, such as selecting, standardizing inputs, adjusting algorithmic parameters, and most importantly, proposing geological entities to explain the different modeling situations (Laouici et al., 2024). These decisions are shaped by the modelers' expertise, prior experience, and domain knowledge (Brodaric et al., 2004). Additionally, due to the focus of current modeling tools on geometric representations, experts frequently make decisions to omit some entities or introduce assumed ones in their geological models to replace or simplify real world geological entities that the systems fail in representing.

Challenges

While it may be expected that the final model would encompass all aspects of related knowledge, in practice only a limited amount is included, largely because it remains tacitly within the minds of experts, especially those about interpretation choices, omission, replacements of entities, simplification, and the iterative adjustment of parameters (Bond et al., 2012; Howard et al., 2009; Kessler et al., 2009; Raab and Frodeman, 2002; Terrington et al., 2008; Wellmann and Caumon, 2018). As a further limitation, final models often consist of 3D mathematical representations with drastically less semantic content compared to the real geological entities subjects of these representations. This further hinders the comprehensive sharing of understanding and insights among experts and systems. Despite significant advances aimed at incorporating more knowledge in geomodelling systems (Grose et al., 2018; Laurent, 2016; Laurent et al., 2016; Wellmann and Caumon, 2018), a significant portion of knowledge used by experts during interpretation remains in their minds. Experts can effectively make interpretations and critical decisions during the model construction, but due to human nature these interpretations and decisions are subjective, biased, and occasionally error prone. Therefore, when most knowledge is tacit it is challenging to account for subjectivity, bias and reasons behind errors. Moreover, it leads to the loss of modeling history and limits the contribution of systems to the interpretation process.

Explicit representation of the tacit knowledge for a geo-model should minimally improve explicability and reproducibility of the model. The knowledge also might be leveraged to guide model construction. To this end, prior work on the use of explicit knowledge for model-building is introduced by Laouici et al. (2024, which corresponds to the second Chapter of this manuscript). Their approach captures expert knowledge in an external model, separating interpretation from spatial representation, and introduces a four-stage interpretation algorithm (Plan, Do, Check, Act). However, the focus is primarily on structuring the interpretation process for automation, with little attention to formalizing knowledge from an ontological perspective.

Proposed solutions

To address these challenges of tacit knowledge loss, knowledge sharing, standardization, and incorporation into geomodelling systems, this chapter proposes the use of conceptual modeling and formal languages via the deployment of ontologies, powerful semantic technologies that provide formal and explicit specifications of a knowledge domain (Brodaric et al., 2008; Gruber, 1995; Guarino et al., 2019; Guarino and Giaretta, 1995; Guizzardi, 2007, 2005). More precisely, this Chapter presents an ontology, named POKIMON (Processes, Observations, Knowledge, Information, and Modeling ONtology), to structure the knowledge used during the geomodelling process with focus on the following general competency questions:

- Which knowledge is critical to building 3D models and their representations?
- How to model this knowledge using formal ontology languages?
- How can the processes employed to build models be represented with formal languages?

Related work

As discussed in details in Chapter 1, the usage of formal languages and ontologies is growing in the field of geosciences (Abel et al., 2012b; Breitman et al., 2007; Brodaric and Gahegan, 2010), with numerous standards and knowledge models developed (Fauziati and Watanabe, 2010; Lombardo et al., 2018; Mantovani et al., 2020; Morandini et al., 2011; Raskin and Pan, 2005; Simons et al., 2006). Many of these are focused on small portions of geoscientific knowledge, such as structures, faults, and the geological time scale (Babaie et al., 2006; Cox and Richard, 2015; Zhong et al., 2009). In contrast, the GSO (Brodaric and Richard, 2020) and GeoCore (Garcia et al., 2020) ontologies represent geoscience entities at a broad level. In the 3D geomodelling domain, ontologies are used to describe metadata, query heterogeneous information (Mastella et al., 2009), describe geomodelling services (Belaid et al., 2011, 2009), harmonize data between geo-models, particularly for fault structures (Qu et al., 2023), and formally represent topological and geometric properties (Wang et al., 2016; Zhan et al., 2022). While these ontologies –dedicated for 3D modeling– describe a range of geological entities in ways similar to general geoscience ontologies, most of them typically do not reuse these domain ontologies, despite the importance of re-use in developing interoperable tools (Fernández-López and Gómez-Pérez, 2002; Gruber, 1995; Katsumi and Grüninger, 2016) and very few align with top level ontologies, hindering convergence with other domains.

An important consideration in representing 3D geo-models is the distinction made between a model, its representation, and visualization (Perrin and Rainaud, 2013). These distinctions are grounded in the definitions provided by Guizzardi (2003) and Guarino et al. (2022), which will be further explored in this paper (section discussion xxx). These definitions are used to develop an ontological framework that organizes the aspects of the visualization component (Abel et al., 2019), namely the visual contents describing geological entities (e.g., images, maps, seismic profiles). However, the model and representation components are not yet addressed ontologically, which inhibits knowing the nature of the interpretations and other assumptions behind any representation or visualization of a model.

Furthermore, although efforts to represent the underlying geological theory is nascent (Perrin and Rainaud, 2013), primarily for entities such as geological and natural laws, principles, and empirical observations, they have not been applied to the 3D geological interpretation process from an ontological perspective.

3.2. Use cases

To define the ontological requirements, five use cases are examined, each addressing distinct challenges encountered in 3D structural modeling. By "ontological requirements," I refer to the key aspects that an ontology must describe and support, expressed through the concepts, relationships, and axioms it must represent. Similar to software requirements, these define what the ontology should accomplish and the capabilities it should possess. Built on the generic previously defined competency questions, the defined fourteen requirements (see below) provide further refinement and guidance for the ontology's conceptualization and formalization.

Use Case 1 focuses on the various types of geomodels according to the concept of representation, while **Use Case 2** explores model preservation and the conditions under which a model is considered different. **Use Case 3** addresses the considerations involved in simplifying models, and the origin of the information behind the representation. **Use Case 4** focuses on the objects and processes involved in interpretation, and **Use Case 5** examines the role of interpolation and its relation to interpretation. Although these use cases are not exhaustive, they highlight critical areas where knowledge is often tacitly applied. As a result, POKIMON is designed to be adaptive, evolving over time to accommodate new knowledge aspects and advancements in the field.

Photographs used in these use cases are taken from a rock quarry in western France, within the Central Armorican Domain (CAD) in Brittany, whose geology was introduced in Chapter 1, section 1.1.3. *The Grès de Chatellier Formation*, addressed in the use cases, a 100-meter-thick layer Ordovician in age, consists of sandstone, siltstone, and quartzite beds. These beds are slightly metamorphosed and lack clear lithological separation (Verhnet, 2010), thus further distinction of sub parts of the units and modeling them are only synthetic to demonstrate the ideas on interpretation and the use of expert's knowledge. Below, we detail the 5 use cases with the derived ontological requirements. For clarity, the requirements are presented in Roman numerals within each use-case and summarized at the end of this section.

Use case 1: geomodels and their representation

Figure 3-1A presents a photograph of a rock quarry. Experts can describe the geological configuration in terms of lithologic and stratigraphic properties of the geological units, without exactly locating this description in space, thus, providing a conceptual model. Experts could also draw these aspects as a cartoonish figure or illustrate them as a stratigraphic column to highlight relations with other units (Figure 3-1B). In fact, experts can build geomodels by instantiating classes conforming to various rules, without necessarily building a full geometric description of the geological objects and structures. The resulting conceptual model often lack spatial properties like exact location or shape but includes fundamental topological relations, such as which unit lies above another or which fault cuts through a unit (Thiele et al., 2016a).

In fact, the conceptual model captures the way that entities fit together, what they are, and possibly how they came to be, without exactly specifying their position in space. 3D modeling tools often provide a dedicated digital geological module to integrate large part of the information in the conceptual model within their mathematical framework as discussed in Chapter 1. Therefore, ontological representation of a **conceptual model** is a key requirement (I) for the proposed POKIMON ontology.

Experts can further develop the conceptual model into a more detailed model by incorporating additional properties, particularly spatial and temporal attributes. For example, the model in Figure 3-1C is geometrically and topologically complete, and it also captures various relationships between objects, which are informed by the underlying conceptual model. It is this model that is represented in a geometric digital representation. Therefore, the ontological representation of a model that includes fully these **spatial and temporal properties** is essential (II). Both type of models could be expressed in different ways, such as in drawings, maps, 3D models, or in texts (Liben et al., 2010). For example, Figure 3-1 shows the study area represented as a photograph (Figure 3-1A), digital 3D model (C), and text (D), each with their own characteristics such as scale, resolution, or type of police. A key requirement is the representation of **representational properties of models** (III)

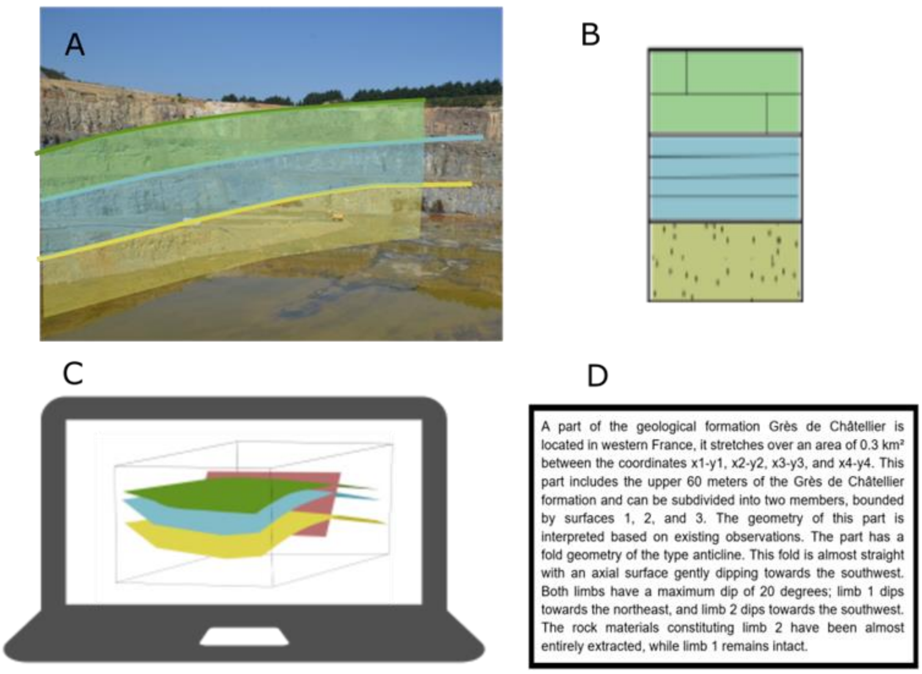


Figure 3-1 Use-case1: representation of the study area (A) and different types of models. The conceptual model about this area is shown in B, and a fully described model is shown in a digital 3D model in C, and as text in D

Use case 2: geomodels integrity

A good example of experts tacitly applying their knowledge is in determining whether a model should be considered the same after modifications as shown in Figure 3-2. For example, a model with folded surfaces (Figure 3-2A) can be slightly modified with a different value of the wavelength of the fold structures (Figure 3-2B), or through adding a secondary fold (Figure 3-2C). Some experts may consider these modified models to be the same as the original, while others may regard them as different. However, this distinction is often made without explicitly recording the rationale or criteria within the system. While such decisions may not be critically important when human experts are interpreting models, they become decisive when systems are tasked with making suggestions or proposing interpretations, as per Laouici et al. (2024). This distinction has direct implications for computational time and memory usage during the modeling process. Therefore, a key requirement for the POKIMON is to define when a model no longer **retains its integrity** after modifications or updates to its components (IV).

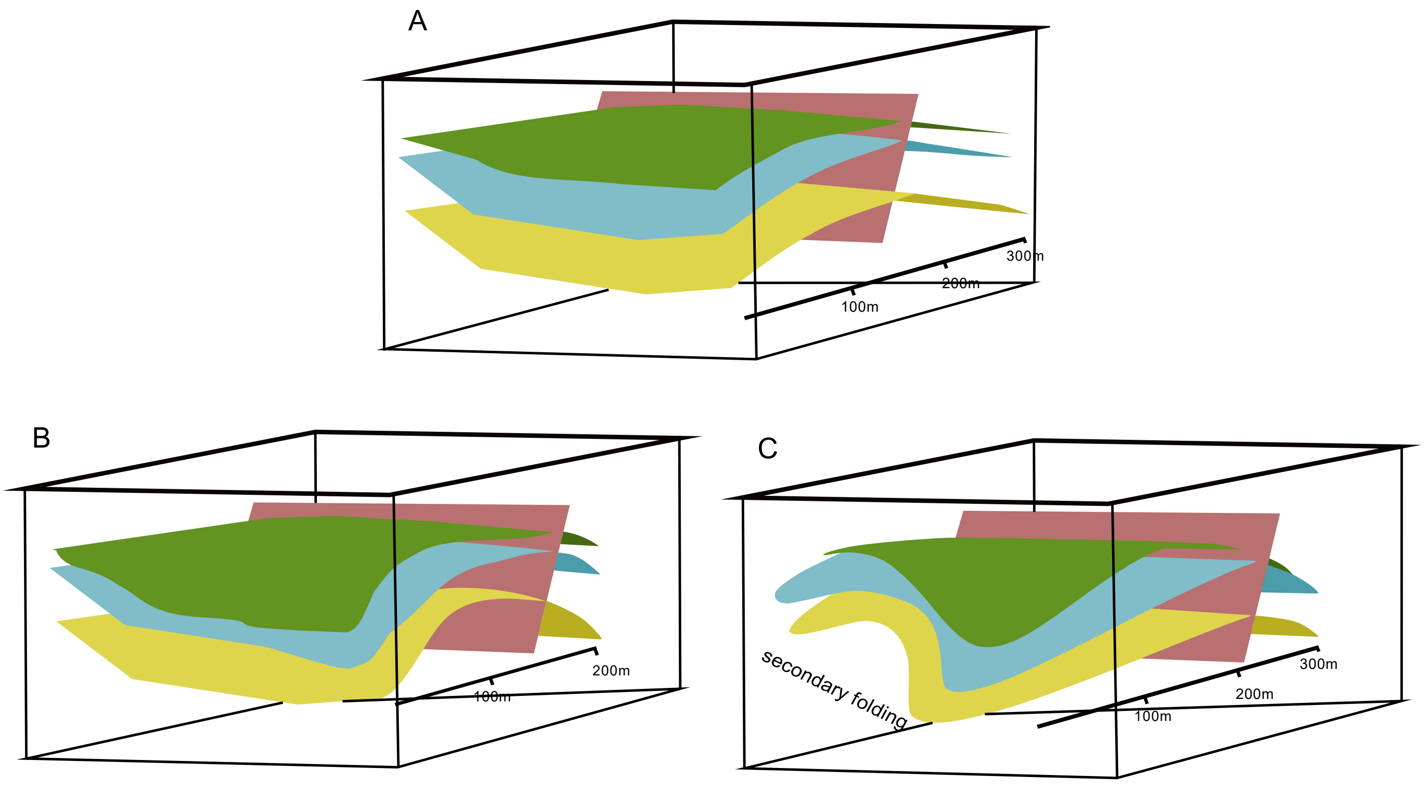


Figure 3-2 Use-case 2: a geological model of three folded surfaces (A) and its modification is in two different ways by reducing the wavelength of the principal fold (B) and by adding a secondary fold (C), evoking the question of whether it is each time the same or different model

Use case 3: simplification in geomodels and origin of entities

While building models, experts often simplify what they see in reality by selectively including only critical entities, and neglect others, typically without a record of these actions. However, such simplification often has different intentions. For example, Figure 3-3A shows a geological unit fractured by a series of faults with very short displacement. Experts would likely include these faults in a reservoir-oriented model (Figure 3-3B), as they may influence fluid flow, but omit them from a tectonic-oriented model (Figure 3-3C) as these structures have less significance on the overall tectonic configuration of such study. To do so, experts often define certain constraints that the geological entities must honor to consider them for modeling. If models about geological entities are proposed during the process of modeling, and they do not honor these constraints they would be omitted from the knowledge base and replaced by others. Sometimes, simplification involves adjusting some properties of geological structures (e.g., the roughness of a surface) or replacement of entities by others only for representation purposes while keeping original semantic descriptions. E.g., Figure 3-4A shows a series of geological strata. The original geomodel as shown in Figure 3-4B represents this configuration with full details. Experts will use this original model of real entities during their studies, but for communication purposes, they could simplify it Figure 3-4C) for a meeting with stakeholders for instance. In practice, modeling systems are not equipped for such manipulation and the solution is often duplicating the model with no links between the simplifying and simplified models. The ontology must then enable tracking the **constraints** (V)**, and the simplification and detailing of** entities (VI).

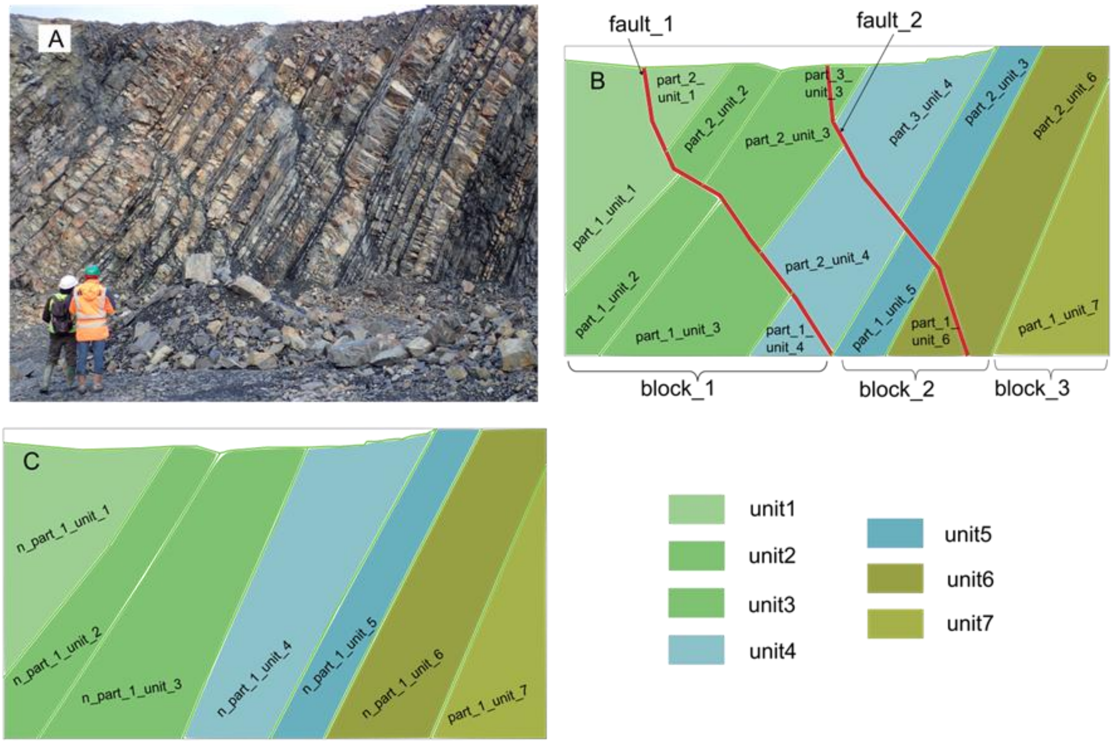


Figure 3-3 Use-case 3: a real fractured outcrop (A) and its possible modeling with all details (B) or with no faults having short displacement (C).

In other cases, experts resort to simplification, but this time because the deployed modeling system is not able to represent the study area with its full semantic content. Thereby, experts introduce entirely new entities that do not exist in reality for representation purposes For example, when modeling a configuration consisting of multiple thrusts with flat ramps and folded layers (Figure 3-5), one solution is to introduce to the geomodelling system a conceptual model that ignores the faults and focuses only on the faulted blocks while considering them as separated units with assumed contacts between them (onlap). This means there is an entire configuration that is assumed and not observed; hence, the ontology must enable tagging the nature of an object, e.g., **observed, interpreted**, or **assumed** entities (VII).

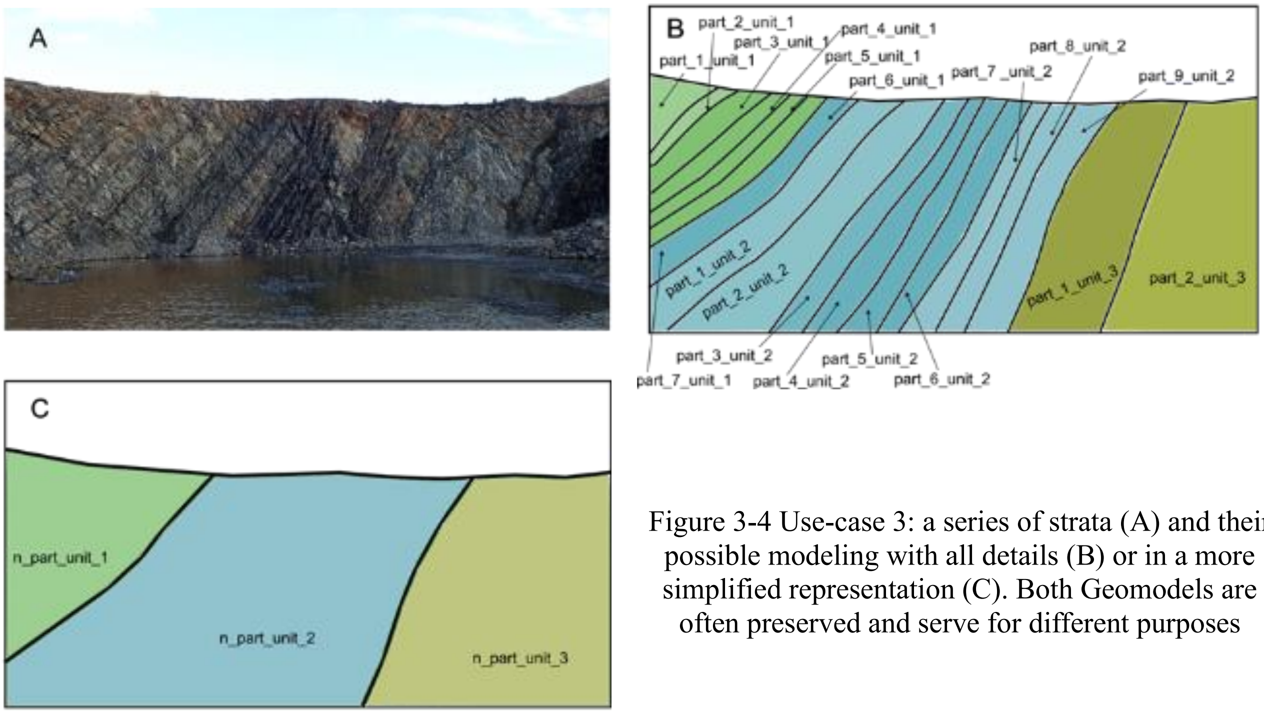


Figure 3-4 Use-case 3: a series of strata (A) and their possible modeling with all details (B) or in a more simplified representation (C). Both Geomodels are often preserved and serve for different purposes

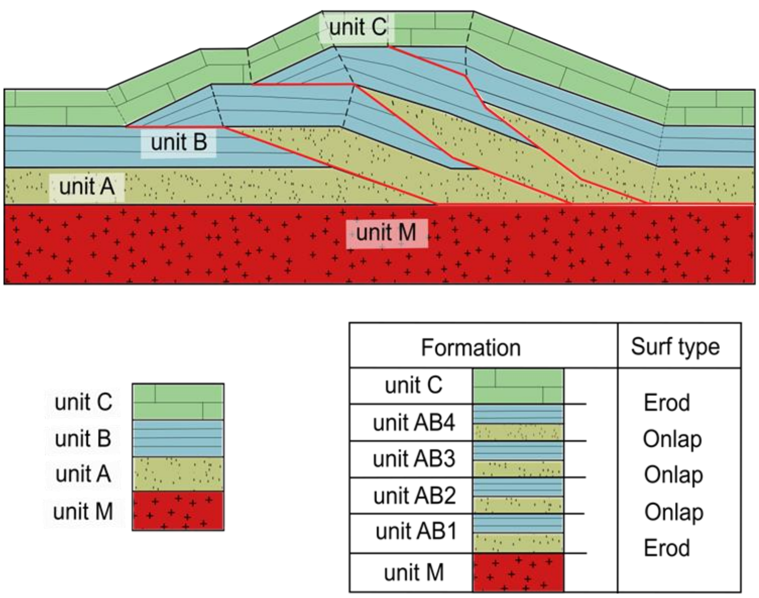


Figure 3-5 Use-case 3: a configuration of thrusted blocks with ramps and folds and their modeling. The model reflecting the real geological configuration is shown in the left lower part of the figure and the model that is introduced to geomodelling systems for representation is shown on the right lower part of the figure (adapted from the stratigraphic pile of the software GDM multilayer*)

*https://www.brgm.fr/fr/logiciel/gdm-suite-gerer-modeliser-visualiser-donnees-geoscientifiques-3-dimensions

Use-case 4: processes in geomodels building

In the proposed use case, geological observations are collected along accessible roads, following a linear path across the outcrop. Constructing a model from these observations requires substantial interpretation of the input data. Figure 3-7 illustrates this geological context, highlighting the recorded observations and measurements:

- As per (Laouici et al., 2024), the interpretation process begins with an unexplained geological situation, such as the observations in Figure 3-6 A. These observations give rise to three possible situations derived from the principles of spatial proximity and lateral continuity (Figure 3-6B). Each situation is an input to the interpretation process, with numerous iterations generating distinct model parts (Figure 3-6C), resulting in a final interpretation and model (Figure 3-6D). Requirements for ontological representation then include: (VIII) representation of **situations** and their links to initial observations.

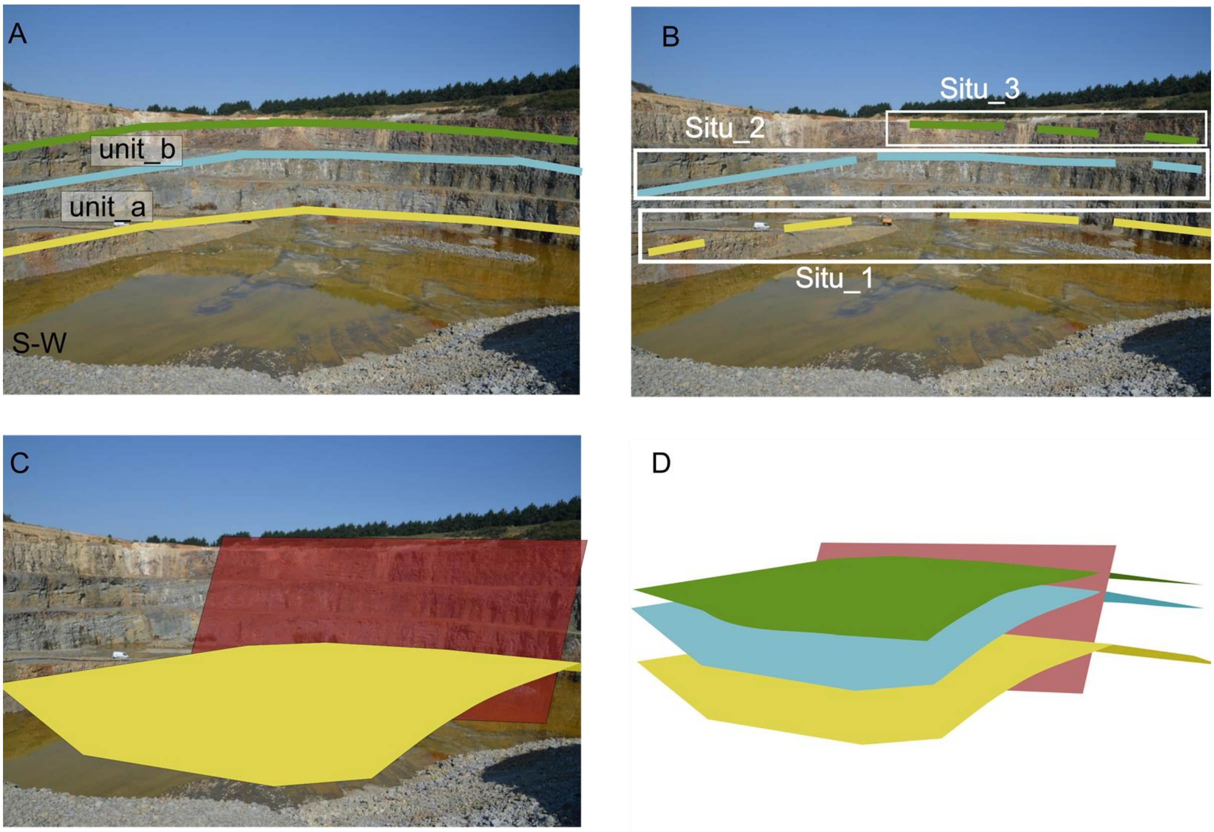


Figure 3-6 Use-case 4 the interpretation process to construct a model from field observations. (A) an outcrop in stone quarry with possible observations collected along accessible road. (B) shows the massaging of these observations and the formulation of some possible situations requiring interpretation according to Laouici et al. (2024). (C) shows a first possible interpretation (a fold, the yellow curved surface, and the red plane represents its axial plane)). (D) shows a final model of the study area after various interpretation iterations. note that these observations and interpretations are synthetic to show the m proceeding of the modeling process from raw inputs to a final model

- As per (Laouici et al., 2024) the interpretation process is general with several subprocesses that might run in different ways. For example, the selection of a situation to interpret could be done randomly, by expert choice, or spatially from bottom to top or left to right. The ontology thus must represent (IX) **interpretations**, and (X) the **various kinds of processes involved in interpretation** as well as links to the interpretations.
- Each interpretation iteration must comply with geological knowledge, including conformance to classes and geological norms. For instance, in this use-case a folded layer is proposed in (Figure 3-6C), which must comply with the Fold class and with other principles such as the principle of original horizontality. Therefore, the ontology needs to include rich **geological object classes** (XI) and **norms** (XII).
- Figure 3-6D shows the resulting model containing 3 folded surfaces, which should be evaluated to be geologically plausible during an evaluation process and then it is for the interpretation to accept or reject the proposed geomodel during a decision process. The ontology must then allow the association of geomodels to the results of **evaluations** and **decisions** (XIII).

Use case 5: interpolation vs interpretation

A final use case that highlights a gray area where experts tacitly apply their knowledge is the precise purpose behind the use of interpolation. Interpolation, which estimates properties at unknown locations based on known values (Franke, 1982), is primarily used by most commercial modeling tools to build the geometric representations of modeled interfaces, such as unit boundaries or fault planes. Interpolation is used as any mathematical operation to deduce the properties based on some mathematical assertions, and used iteratively by experts to refine geometries by leveraging large datasets and spatial correlations (Marinoni, 2003; Oliver and Webster, 1990), or even to discern potential deformation patterns based on the interpolation results (Laurent et al., 2016).

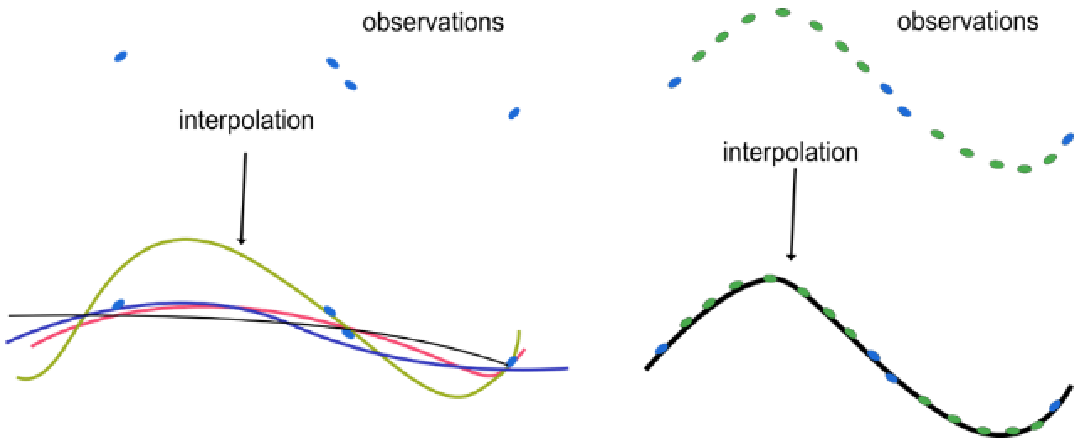


Figure 3-7 Use-case 5: comparison of the role of interpolation. When using a restricted number of observations, interpolation results vary significantly, and the interpolation is used as a tool to estimate the value of the modeled surface parameters and eventually inferring associated structures. On the other hand, when there is exhaustive coverage, interpolation is used only to produce the representation for visualization as the objects, structures, and their parameters are known

However, when the geometry is already known through previous interpretations, or simply through observation, the interpolation is only applied here to reconstruct the geometry of the model that will be represented visually and does not interfere with the estimation of its properties as sketched in Figure 3-7. Thus, the ontology must also provide a framework to capture when **interpolation** (XIV) is used to estimate values and help detecting new concepts (as part of the interpretation process) or just used in enhancing known geometries and reconstructing them.

In summary, the key ontological requirements for POKIMON design are as follows:

- (I) Ontological representation of conceptual geological models
- (II) Ontological representation of geologic models with complete spaciotemporal properties
- (III) Ontological representation of the representation properties of geological models
- (IV) Criteria and properties related to model's integrity
- (V) The notion of constraints when building geological models
- (VI) The simplification and detailing aspects
- (VII) Origin of entities described by the model (observed, interpreted, assumed)
- (VIII) Situations considered during interpretation
- (IX) Interpretation and the history of interpretation.
- (X) Processes associated to the interpretation and their specifications
- (XI) Geological entities
- (XII) Geological norms
- (XIII) Evaluation and decision results and their relation to geological models
- (XIV) Interpolation usage and purpose behind

3.3.Ontology Design Methods

As outlined by the use cases, POKIMON has to deal at least with two overlapping domains: (1) the geomodelling domain, including entities involved building a geological model (e.g., processes, representations), and (2) the geological domain, including the various natural entities that are modeled. Accordingly, POKIMON is built by reusing and extending selected components from several domain and top-level ontologies: (1) entities from the Basic Formal Ontology (BFO) (Otte et al., 2022), the Information Artifact Ontology (IAO) (Ceusters and Smith, 2015), and the Information Ontology Foundry Core Ontology (IOF) (Drobnjakovic et al., 2022; Karray et al., 2021), and (3) the Geoscience Ontology (GSO) (Brodaric and Richard, 2020). Components from BFO, IAO, and IOF are used to address the knowledge aspects in the geomodelling domain, specifically concerning the representation of information, while components from GSO are used to address the ontological representation of the natural entities. In doing so, POKIMON follows an extraction and extension approach, as per Katsumi and Grüninger (2016), in which only certain axioms are extracted from the original ontologies and new axioms are introduced, such that the basic intent of the original components remains intact.

As many of these original ontologies are quite large and diverse, and conceptually misaligned in areas not significant to POKIMON, this approach is adopted mainly to simplify ontology construction, to avoid conceptual and logical conflicts, and to ease reasoning and ultimately achieve logical consistency. In essence, POKIMON requires a few key components from the selected ontologies, so a full import of all their contents is unnecessary, and this allows POKIMON to remain compact, internally consistent, and easily debugged. To better understand how POKIMON builds on these ontologies, main components extracted from them are explained next and summarized in Figure 3-8.

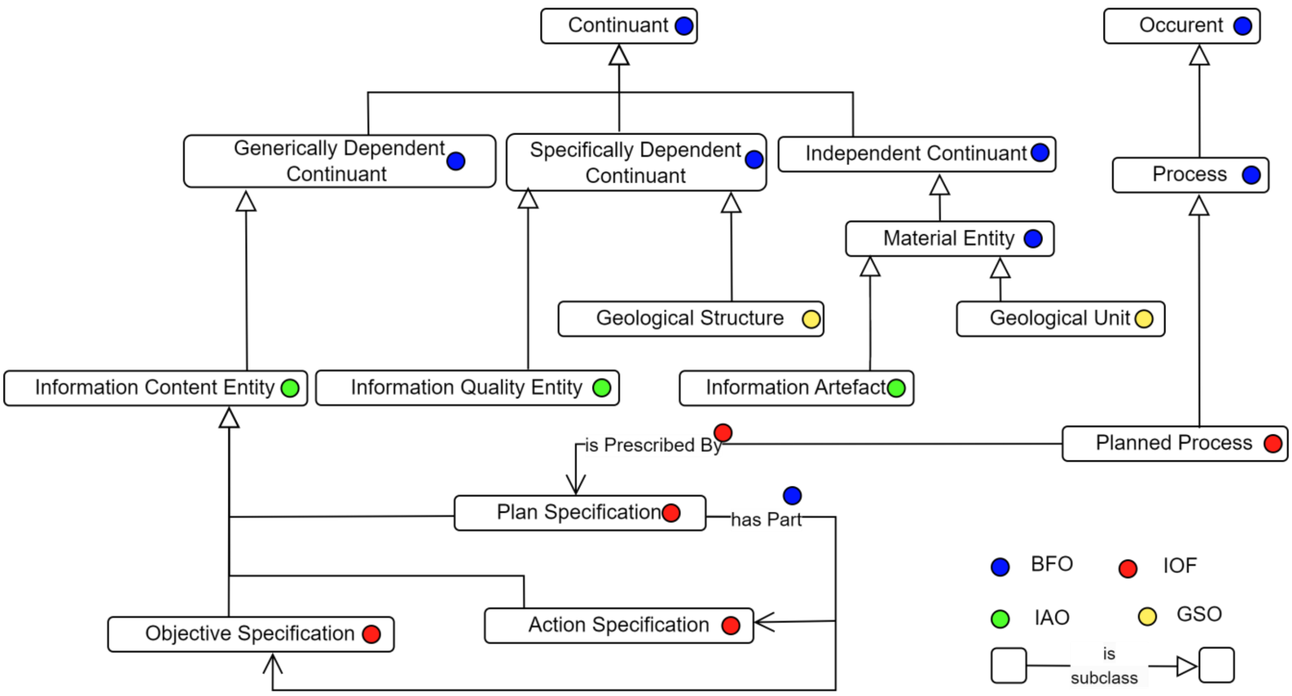


Figure 3-8 Schematic diagram showing the main entities borrowed from BFO, IAO, IOF, and GSO. These entities represent the skeleton of POKIMON

BFO and its components

The *Continuant* and *Occurrent* classes are key distinctions adopted from **BFO**. Continuants are fully present at any timepoint, as they persist in time (e.g. a rock), and occurrents are never fully present at a timepoint, as they unfold over time (e.g. an earthquake). Continuants are classified into three types, delineated by their ontological dependence on external entities; external entities are not parts, qualities, or constituents of the dependent, nor the processes it participates in. Then (1) a *Generically Dependent Continuant* exists due to some instance of a type existing, e.g. the content of a mineralogy book is generically dependent on the hardcopy of the book as it can be represented in different copies; (2) a *Specifically Dependent Continuant* exists only if another specific entity exists, e.g. the color of an amount of mineral exists only if the specific amount of mineral exists; and (3) an *Independent Continuant* exists independent of any external entity, e.g. an amount of mineral is not specifically dependent for its existence on any other external entity. An important relation is used which *inheres in* –and its *inverse bearer of*– is used to line a specifically dependent continuant to the entity it inheres in it.

IAO and its components

IAO extensions of BFO relevant to POKIMON include those components that enable representation of entities such as documents, databases, and digital images, as well as their contained information. In doing so, IAO delineates the information present in a representation from the carrier of the information, so the information can be carried by multiple representations, e.g. a geological model can be carried in this sense by the lego blocks on my desk, a certain pdf file on my laptop, or in my mind. The information multiply represented is an *Information Content Entity* (ICE), which is abstract (it does not extend in space in time), and the I*nformation Bearer Entity (IBE),* that becomes an *Information Artefact* (IA) when the ICE is concretized through *Information Quality Entity* (IQE) (Smith et al., 2013), such that the IQE embeds the ICE in the IA. For example, if a book is made of a series of abstract symbols, then those symbols are embedded in a physical copy of a book by being inherent in ink patterns on the book's pages; or they are inherent in the bits of a digital file that make up the file (Otte et al., 2022) as shown in Figure 3-9A. These elements are related in the following way: the ICE *is about* something, e.g., a book on geology is about geology; the IQE concretizes the ICE; and the ICE is generically dependent on the IA, with the IQE specifically dependent on the same IA – thus neither the ICE nor IQE can exist without an IA.

IOF and its components

IOF (Drobnjakovic et al., 2022; Karray et al., 2021) extends both BFO and IAO, and is a core ontology for data, processes, and information interoperability in manufacturing. Of particular interest is *Planned Process* which is an occurrent that runs over time following the steps laid out in some abstract specification, its ICE, as concretized in an IQE and manifest in an IA such as a hardcopy operations manual or digital software code. As per Figure 3-9B, the plan specification has two ICEs as parts, an action specification and an objective specification, which respectively describe the steps to be run and the purpose of the plan. In a variation of the *is about* relation, the process *is prescribed by* the plan specification. Such plans are relevant to POKIMON, inasmuch geological modelling and interpretation are planned processes.

GSO and its components

GSO has three layers. Its top layer includes general classes applicable to any discipline, adapting elements from top level ontologies such as BFO, DOLCE (Borgo et al., 2022), and UFO (Guizzardi and Wagner, 2010). The second layer contains geoscience classes; this layer is meant to cover the range of entities in geology, such as geological objects, materials, structures, qualities, geologic time, and geologic relations. The third layer extends the general geoscience layer into specific geologic subdomains, for example for geologic units or structures, drawing on aspects of data exchange standards such as GeoSciML (Simons et al., 2006). GSO is a comprehensive ontology including over 8000 entities, but POKIMON utilizes only a few GSO entities, primarily the classes of: Fold, Fault, Geological Unit, and Contact. Mapping these entities in POKIMON will be further detailed in section 3.4.1.b.

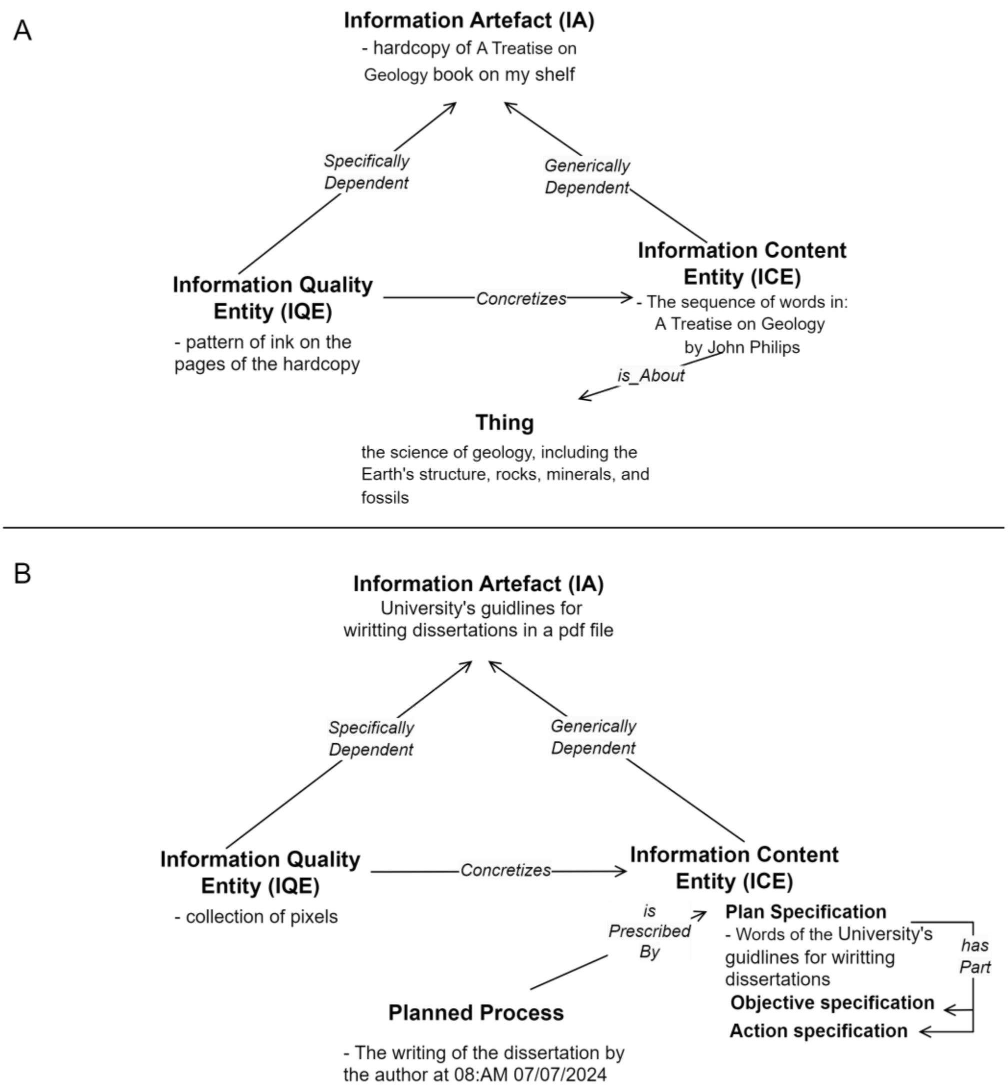


Figure 3-9 Schematic diagram showing the theoretical framework of IAO and IOF. Panel A describes a physical book as an "Information Artefact" (IA), specifically, a hard copy of A Treatise on Geology. The Information Quality Entity (IQE) is the pattern of ink on the pages of the hardcopy, which "concretizes" the Information Content Entity (ICE), the sequence of words within the book. This ICE is generically dependent on the IA and is about a "Thing", the science of geology, encompassing Earth’s structure, rocks, minerals, and fossils. Panel B describes a Planned Process 'the writing of the current dissertation by the author at 08:00 AM on 07/07/2024’. This process is prescribed by an ICE, which consists of the plan specification for writing dissertations provided by the university. The ICE is concretized by an IQE, a collection of pixels on the author’s laptop

3.4.The POKIMON Ontology

At its current version, POKIMON consists of 180 classes and 65 object properties, including both proposed entities, and those borrowed from the employed ontologies. These entities serve as the foundation for modeling the key concepts and relation central to POKIMON's domain of application. This section details the theory, developments and extensions made within POKIMON, focusing on two main aspects: models and processes. The first subsection, "Models in POKIMON" addresses how different types of models are represented and characterized within the ontology using the adopted framework from existing ontologies (requirements I to VIII, X and XI defined above). The second subsection, "Processes in POKIMON," explores the formalization of the geomodelling processes with focus on interpretation according (requirements IX, XII and XIII), thus tackling the dynamic behaviors and interactions between entities within the ontology. The proposed ontology entities with their brief definitions are summarized at the end of this section.

3.4.1 Models in POKIMON

POKIMON fundamentally distinguishes objects and relations in the real world from their analogs in the modeling –representation– space, such that a representational entity is about something in reality. This representational entity is the *model* –or *geomodel* when it is about geological entities– which is further delineated into abstract and depictional components, in that any abstract component of a *geomodel* can be depicted in multiple ways, such as text in a document, a 2D cross-section, or a 3D digital file. The abstract component is modeled in POKIMON by the class *Semantic Content Entity* related to geological entities by the relation *has Semantic Component* as shown in Figure 3-10. The geological entities are modeled ontologically via a geology domain ontology (see next section for further details). The *Semantic Content Entity* is abstract in the sense that it is only existing as an ontological representation because the semantics ought to be captured and modeled to enable their depiction in various external formats. This ontological representation will function as an internal model within the geomodelling system, analogous to how humans mentally represent geological concepts.

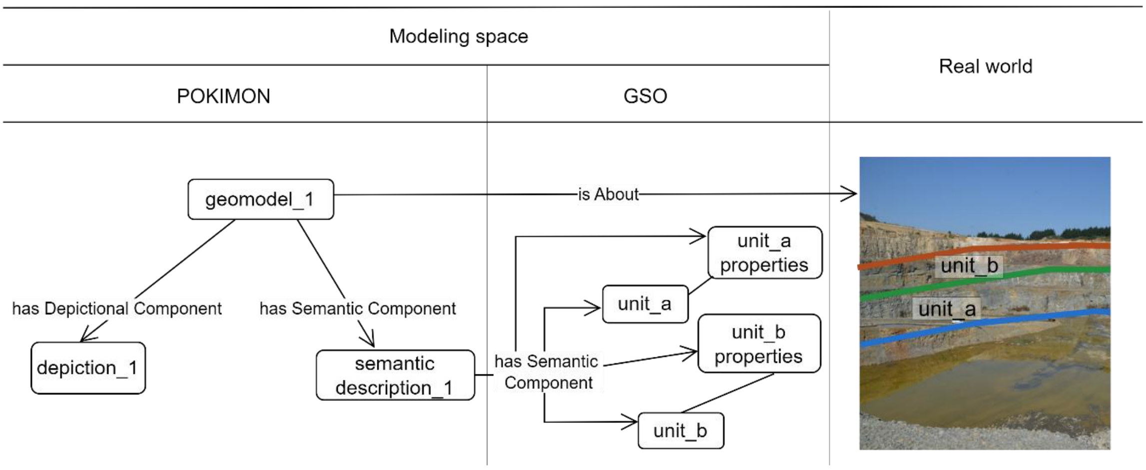


Figure 3-10 Schematic diagram depicting an instance of the type geomodel about a real-world geological configuration. The geomodel_1 has two components, depiction_1 instance of the depiction content entity, and sematnic_description_1 for the semantic content entity. The semantic description describes two geological units with their properties using the GSO mapping.

When such models are depicted, they are externalized, either by humans as in speaking or writing, or by the system as in a 3D figure. Following this, the depiction component is modeled in POKIMON using a *Depiction Content Entity*. This latter is a casting of the semantic component into a specific external representational paradigm, e.g., as a sequence of abstract symbols that make up a textual description, or a figurative configuration of iconic symbols that make up a diagram. It is this depictional representation that is embedded in a specific IQE and thus IA, such as in a specific copy of a book, map, or digital model. Next, we define the main entities used in this formalism as shown in Figure 3-11.

- *Geomodel*: is an ICE that points to the real geological entities via the relation *is About*. The geomodel has two components (1) a Semantic Component which is the *Semantic Content Entity* (SCE), and (2) a depictional component *Depiction Content Entity (DCE)*, in a sense, the same semantic content can be depicted in different ways through various DCEs. When it is depicted, the *Geomodel* and its components are concretized by the same IQE, on a given artefact (IA).
- *Semantic Content Entity* (SCE): is an ICE that provides an abstract description of the geological entities. It points via the relation *has Sementic Content* to the different geological entities modeled via a geology domain ontology.
- *Depiction Content Entity* (DCE): is an ICE concretized by the same IQE that concretizes the associated geomodel. The DCE describes the depictional abstract contents making up a representation of the geomodel. Three types of DCE are distinguished:
    - *Figurative Content Entity*: the geomodel is represented using a figure. The figure can be a volumetric figure such as a 3D geometric model using explicit or implicit surfaces. It can be also a planar entity, such as a sketch, image, section, or transect.
    - *Proposition Content Entity*: the geomodel is represented using a sequence of words or another linguistic construct. This class includes things such as reports, papers, emails, and map sidebars.
    - *Connectionist Content Entity*: the geomodel is represented through a connectionist pattern such as a machine-learning model.
- *is About*: A relation linking the model within the ontology to real geological objects. Since the model cannot directly interact with real objects, this relation remains conceptual and cannot be implemented per se.
- *has Semantic Component:* A relation connecting a geomodel to its *Semantic Content Entity* (SCE) or linking the SCE itslef to entities in the geology domain ontology that describe real geological objects.
- *has Depiction Component:* A relation linking a geomodel to its *Depiction Content Entity* (DCE) and associating this DCE with various representational entities, such as surfaces representing geological boundaries or words in a geological report.

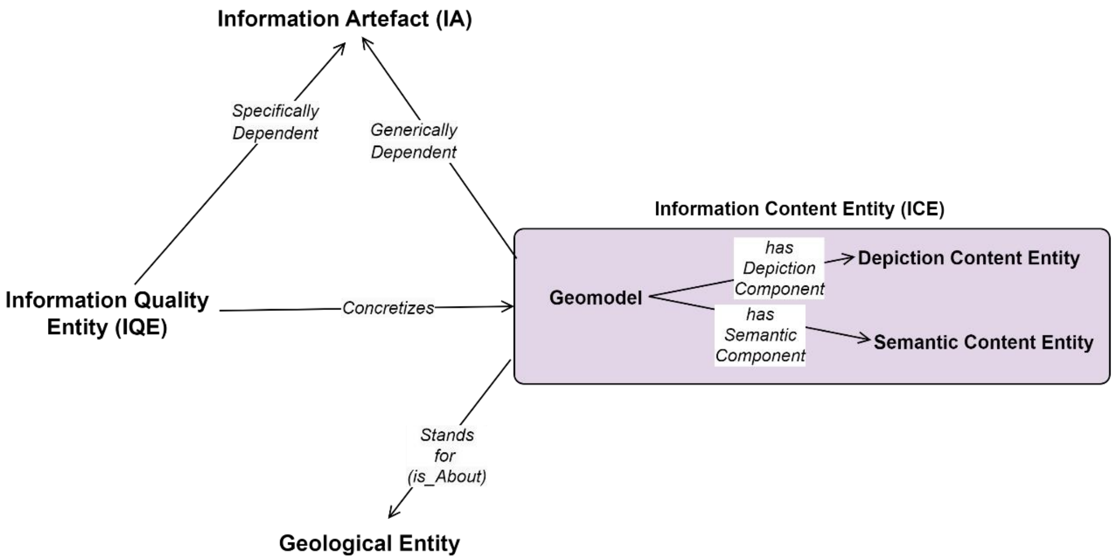


Figure 3-11 Schematic diagram showing the modeling of the geomodel in POKIMON using the IAO framework. The geomodel is a subtype of an ICE, that is about some geological entity, concretized by and IQE and Generically dependent on an IA. The geomodel has a depiction and semantic components both subtypes of an ICE

3.4.1.a Types of Geomodel and properties of geological entities

The *Geomodel* in POKIMON is further delineated into *Conceptual* and *Space-Time Geomodel*. The **Conceptual Geomodel** is focused on defining the concepts at stake and some of their properties for the purpose of helping to discuss or illustrate these concepts, but without a complete characterization of its properties for genericity or simplification. A lot of information is already carried by the description of the structural objects at stake and their topological or relative temporal relations. For example, considering a stratigraphic layer affected by a fault already discusses very important aspects of a model, including topology and relative time, without considering the geometries, angles, and most importantly, the precise current location. Therefore, by definition, geological entities in a conceptual geomodel in POKIMON lack a complete specification of spatiotemporal properties (Figure 3-12). However, these entities could have a relative geographic localization, as being about a region in southern France, or related to a specific entity, such as a fold inherent in a specific geological unit.

The *conceptual geomodel* in POKIMON encompasses the broad "conceptual" model that is introduced to 3D modeling tools which includes namely the different geological units, faults, and their structural relations .

On the other hand**,** the **Space-Time Geomodel**: contains entities that all have precise space time position (Figure 3-12). For example, a geological map describes a specific geological configuration of rocks at a particular location of the Earth's surface and at certain geologic times. Each Space-Time geomodel *satisfies* a conceptual geomodel: behind every space-time there must be a conceptual model. Such satisfaction is asserted by the *satisfies* relation (space-time_geomodel_x *satisfies* conceptual_geomodel_x). POKIMON supports the idea that a single conceptual geomodel can be satisfied by multiple space-time geomodels, reflecting the various realizations possible in the geological modeling process.

It is essential to note that within a conceptual geomodel, geological structures can be recognized as existing without necessarily being instantiated (e.g., "Unit A is folded"). Alternatively, they may be identified as instances (e.g., specific folds or faults) but still lack exact space-time coordinates. On the other hand, within a space-time geomodel, material entities (principally geological units) are considered as parts of the larger units defined in the conceptual geomodel. Any specific spatiotemporal description of a geological unit within the space-time geomodel refers only to a part of the unit, as the unit itself is fully defined only at the conceptual level. This is because geological units typically extend over large areas, often exceeding the boundaries established in structural models. On the other hand, geological structures in the space-time geomodel are described as instances with precisely defined spatiotemporal coordinates, either through newly instantiating structures, or filling space-time properties of already instantiated instances in a conceptual geomodel (Figure 3-12).

As discussed previously; to describe geological entities in POKIMON a set of entities is borrowed from the GSO ontology (see appendix B). Principally, material Continuant class in GSO is mapped as independent continuant material in POKIMON. Geological units are subclasses of this independent material continuant class. On the other hand, geological structures are modeled as features in GSO and could be related to other entities by specific or generic dependency object property. In POKIMON, geological structures are modeled as specifically dependent continuants that depend on the hosting entities.

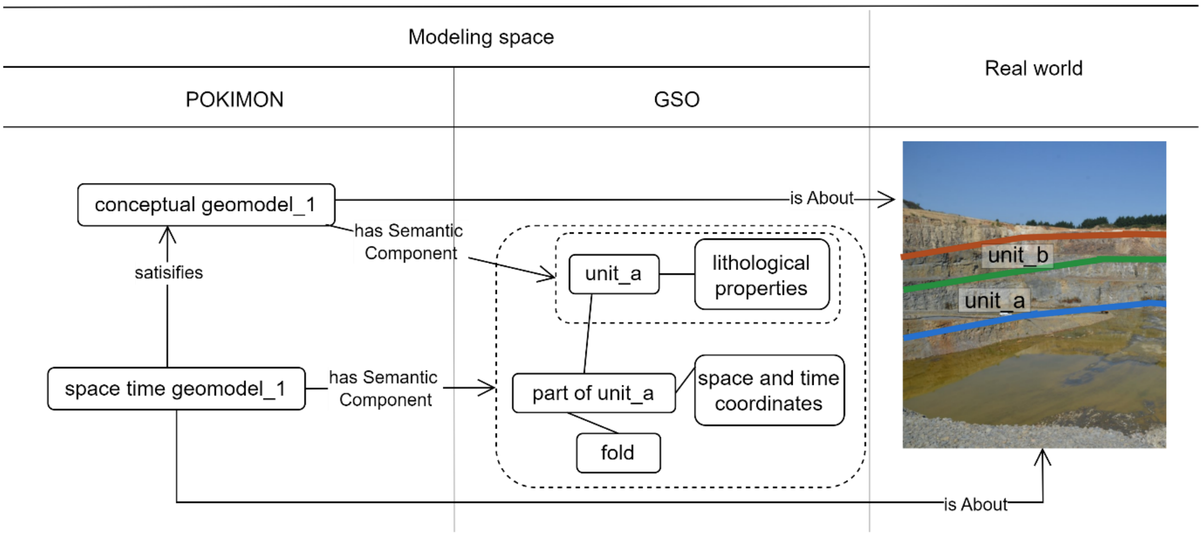


Figure 3-12 Schematic diagram detailing the delineation of the Geomodel into conceptual and space time geomodels

3.4.1.b Observations and origin of information geomodels

POKIMON focuses on the semantics of information derived from observations and measurements, applying the same approach it uses for geomodels by modeling the inherent information within these entities. An observation, at its core, consists of information of a value of a quality that is sensed along real-world geological phenomena during an observation process (Compton et al., 2012). Accordingly, POKIMON addresses observations from the perspective of the origin of the information, recognizing that each piece of information contains a *Semantic Content Entity* (SCE) that is observed. This SCE can then be linked to a *Depiction Content*

Entity (DCE) and concretized via an IQE, such as when an observation is recorded in a field notebook. In this sense, an observation inherently aligns with the concept of a geomodel, although by convention, the term geomodel is typically reserved for a more complex and interrelated set of information. However, by definition, a geomodel is not constrained by the size of objects or the amount of information it encompasses.

During a geomodel construction, the semantic description of geological entities can emerge through direct observation of natural objects, interpretative processes, or assumptions made for representational purposes. Accordingly, POKIMON classifies the origins of geological entities into three types –observed, interpreted, and assumed– and models these origins as qualities, as defined below. The distinction among these origins lies in the degree of certainty regarding the existence of the geological entities being modeled:

- *Observed origin*: A quality assigned when the information is certain, as in cases where it is directly recorded from real-world phenomena in the field.
- *Assumed origin*: A quality indicating that the semantic entity is created purely for modeling purposes, with absolute certainty that such a configuration does not exist in reality. For example, when teaching in the field, a professor may be certain about the geology of a terrain in front of them –such as a series of sedimentary layers– but might ask students to imagine the presence of specific intrusive structures for illustrative purposes. Such imagined geomodel can be then drawn on a piece of paper.
- *Interpreted origin*: A quality applied when entities exist within a probabilistic framework, meaning their existence is uncertain and subject to validation or refutation over time—such as when interpreting seismic data.

3.4.1.c Situations for interpretation

A final geomodel is typically constructed iteratively, with each iteration introducing small portions through the process of interpretation, as illustrated in use case 4. In each iteration, a specific interpretation situation is selected. In geomodelling terms, an interpretation situation refers to a set of geological information that, depending on its origin within the geomodel, either lacks certain attributes necessary for a complete geomodelling representation (e.g., a fault without defined spatial properties) or deviates from established geological norms (e.g., a normal fault exhibiting reverse displacement). Interpreting these situations involves hypothesizing new geological entities –whether objects, relations, or properties– or modifying existing entities to ensure geological coherence. In ontological terms, POKIMON represents a *situation* as an *Independent Continuant* that includes *Semantic Content Entities* as its parts. When a situation is selected for an interpretation process, it acquires a disposition of being interpretable. If the proposed interpreted geomodel succeeds in explaining a given situation, the *explains* relation is asserted between the geomodel and the situation. This relation exists solely at the modeling level, as information entities can explain one another. In contrast, the geological do not inherently possess an "explains" relation but are instead connected through fundamental geological relations (e.g., parthood, spatial continuity, stratigraphic superposition). This distinction underlines the separation between the abstract, explanatory role of geomodels and the physical reality of the geological entities they describe.

For a geomodel to successfully explain a given situation, it must not only provide a valid explanation but also conform to various forms of existing geological knowledge and prior information, as outlined in the following sections.

3.4.1.d Norms and constraints for geomodels

In POKIMON, the geological knowledge and prior information that the geomodels must conform to are norms and constraints. These two entities are information entities that manifest in the world, but do not require to be represented to exist. POKIMON includes the following classes for these entities:

- *Information entity* is an abstract continuant that does not need to be represented to exist, for example natural laws, mathematical, or scientific theories.
- *Norm* is a subclass of an information entity that governs the behavior of other entities. In geology norms include laws (e.g., Walter's law) (López, 2015), principles (e.g., principle of superposition) (De Freitas, 2009), and theories (e.g., plate tectonic theory) (Falvey, 1974) and govern the behavior of natural entities.
- *Constraint* is a subclass of an information entity that defines constraining criteria for the geomodel to determine the pertinence of geological entities considered in the modeling process. At the current stage of study, the constraint entity is used only to set maximum and minimum values of spatial qualities (e.g., length, thickness, wavelength) of geological entities.

Norms and constraints are often known and set by experts at the beginning of the modeling process while geomodels are dynamically created through various iterations of interpretation. Assessing the conformance of a proposed geomodel is basically about checking whether it satisfies the set norms and constraints, the types of entities it describes, and provides a valid explanation for the situation it was selected for as shown in Figure 3-14:

- Ensuring that the proposed geomodel provides a valid explanation for the situation for which it was proposed. For instance, if a geomodel is proposed to explain a geomodel containing a collection of stratigraphic parts, the geometry of the newly modeled stratigraphic part must be consistent with the geometry of the smaller parts, otherwise, the resulting geomodel *contradicts* the initial one (Figure 3-13A and B).
- Ensuring that the proposed geomodel satisfies the set constraints, for instance, the constraint could be the value of displacement along a fault to be considered valid in the geomodelling context. If the proposed geomodel contains faults with less displacement it contradicts the constraints (Figure 3-13C).
- Ensuring that the subject geological entities of a geomodel conform to the necessary conditions of the instantiated types. For example, the geomodel in (Figure 3-13D) is rejected because it describes normal fault concept but the proposed entities exhibit an inverse movement, thus the aboutenness and explanation relations are not verified as a geomodel about a given fault must exhibit the specific movement of its type.
- Ensuring that the geomodel satisfies the geological norms: for instance, a fault cannot cut a younger unit following the cross-cutting principle.

3.4.1.e Evaluations and decisions about geomodels

The assessment of the model validity is basically performed in an evaluation process that occurs during each interpretation iteration as per Laouici et al. (2024) or that could be performed by experts at any point during the modeling to ensure the general consistency of their geomodel in human based-interpretation, mentally or using the geomodelling systems (Parquer et al., 2024).

How the evaluation process is fully modeled in POKIMON will be further detailed (section Processes in POKIMON), at this point the focus is given only to the assertion of the relations of explanation and satisfaction based on the evaluation outcomes. When the geomodel is conform to a given information entity, POKIMON models this conformance via a satisfaction relation (*geomodel x satisfies information entity y*) (Figure 3-13E). If the geomodel is not conform, the relation *contradicts* is used as shown in Figure 3-13C. In addition, POKIMON models the entities resulting from the evaluation process as qualities that *inhere in* the evaluated geomodel. In POKIMON, the quality (called Evaluation) resulting from an evaluation process could have relative (e.g., good/bad), or quantitative (e.g., a probability) values.

The decision of fate of a geomodel (e.g., to be accepted, to be rejected) in the geomodelling process is based on the evaluation outcomes. The decision as a process will be also discussed in section Processes in POKIMON, the important aspect here is that the outcomes of a decision process (called Decision) are considered in POKIMON as a subclass of the realizable entity, a subclass of specific dependent continuant qualities. The Decision class has a relation of specific dependency on the subject being evaluated and also related to it by an aboutness relation. In contrast to IAO where the aboutness relation has only ICEs as domain, POKIMON also employs the aboutness relation between the decision that is an information entity and its subject (decision X *is about* geomodel A).

Finally, if the decision to keep the geomodel is favorable, the *explains* relation is asserted between the geomodel and the situation for which it was proposed. Indeed, if further information comes into play and the geomodel contradicts it, the explanation relation can be negated (omitted from the knowledge base)

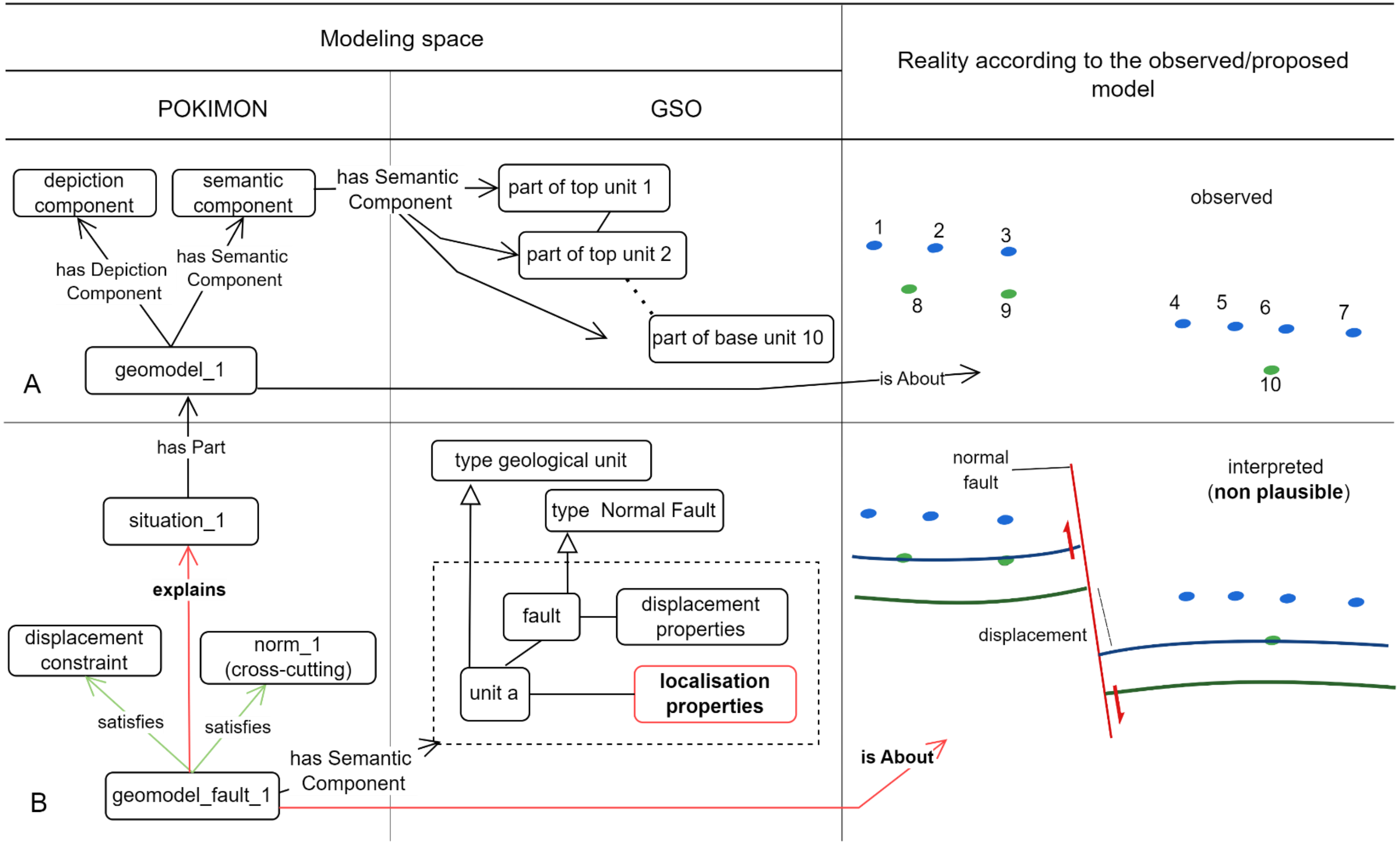
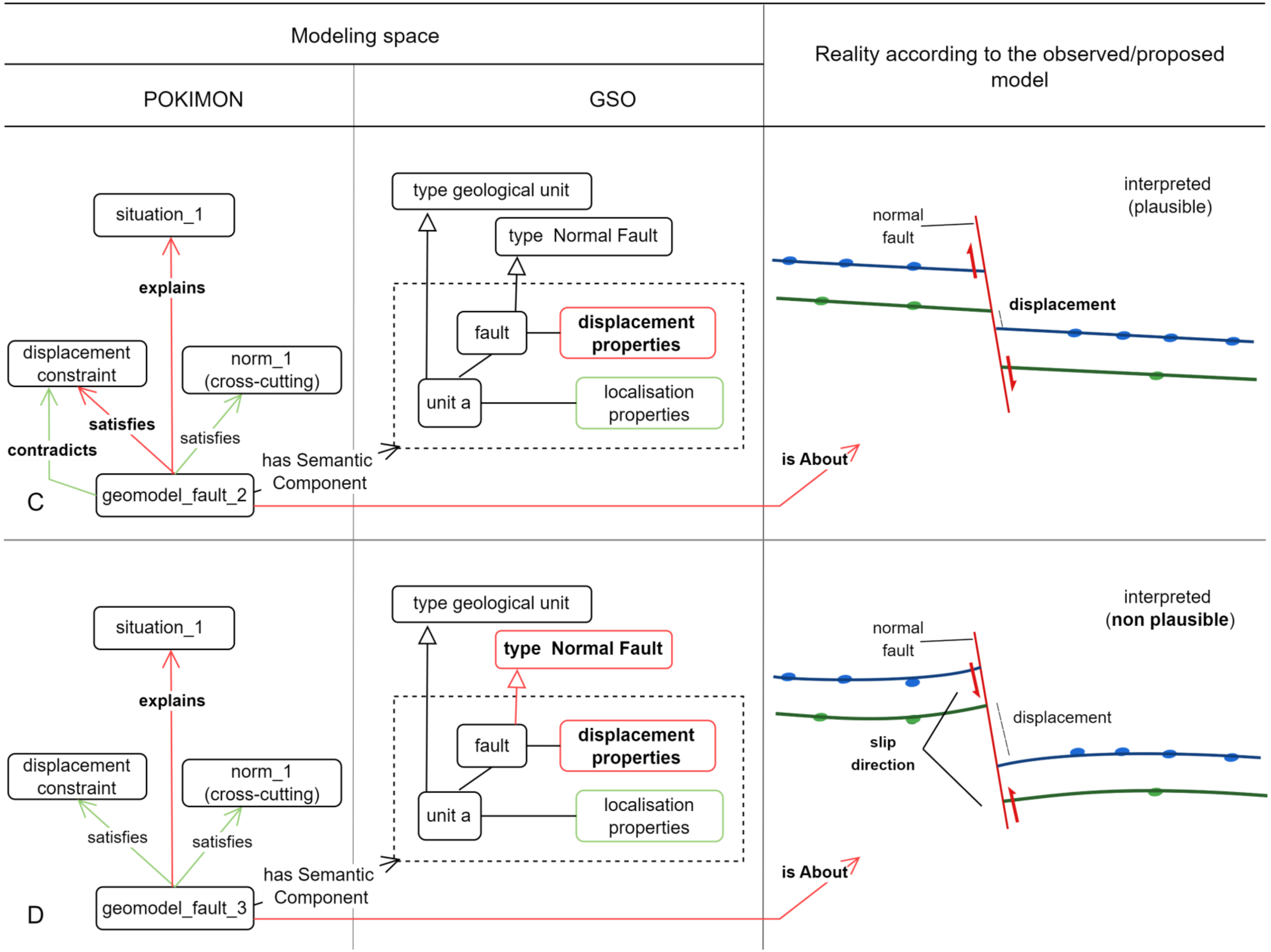
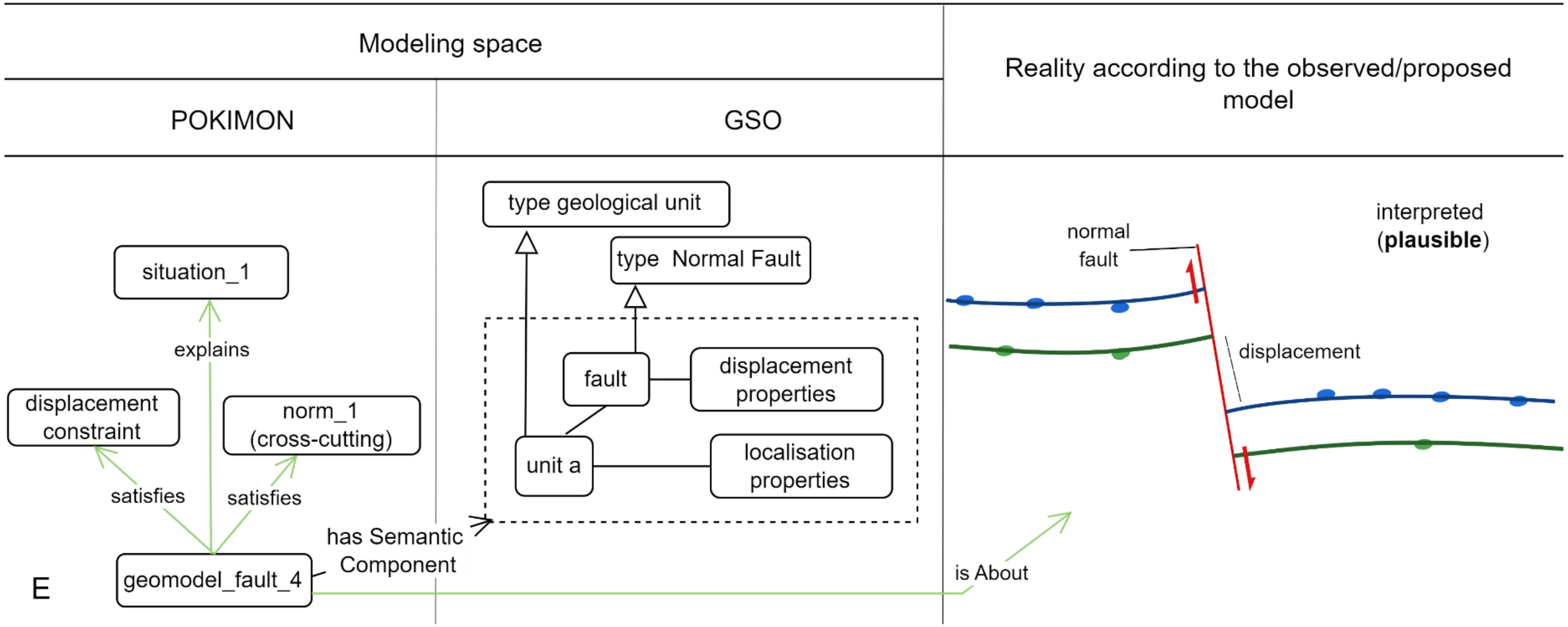


Figure 3-13 Schematic diagram to illustrate a series of geomodels resulting from interpretation and how they relate to information entities and the interpretation situation. Panel A displays some observation from the physical world, their modeling in POKIMON results in geomodel_1. In panel B, Situation_1 is an interpretation situation that has a part geomodel_1. Geomodel_fault_1 which is about a faulting structure is proposed in purpose of providing an interpretation for situation_1. The proposed geomodel satisfies all information entities, however, it is misplaced in the space, thereby the explanation relation for situation_1 is not asserted. (see second part of the figure in the next page)



Part 2 of Figure 3-13. Panel (C) displays a second geomodel (geomodel_fault_2) to explain situation_1. The proposed geomodel satisfies norms, however, the proposed faulting structure has a short displacement contradicting the set constraints, thereby the explanation relation is not asserted. Panel (D) displays geomodel_fault_3 intended to explain situation_1. The geomodel satisfy all information entities, through the proposed structure shows an inverse slip movement, thus contradicting their type (normal fault), thereby the explanation relation is not asserted. (see third part of the figure in the next page)



Part 3 of Figure 3-13. Panel (E) displays geomodel_fault_4 that satisfy all information entities, the type of the proposed structure, and well placed in the space, thereby the explanation and aboutness are asserted

3.4.1.f Characteristics of geomodels

In the POKIMON, geomodels are not only defined and structured, but they are also characterized by a set of key properties that influence how they should be represented and manipulated within systems. The following subsections detail the essential characteristics that define geomodels in POKIMON, including their integrity, their relations with other geomodels, the role of a reference model, and the properties propagated to them from related entities.

Geomodel integrity

During the construction of a geomodel, it is common for it to undergo updates, modifications, enrichments, or deletions of certain components as discussed in the use-case 2**.** A key concept in this process is geomodel integrity, which refers to the conditions under which a geomodel is considered the same or when it becomes a distinct model. In the development of POKIMON, the guiding principle is that a geomodel retains its identity as long as the types of geological entities (e.g., faults, horizons) remain unchanged, even if their specific properties are modified. For example, a model representing a normal fault remains the same if values of properties such as slip direction are adjusted, but it would be considered a different model if the fault type is altered (e.g., from a normal fault to a reverse fault). Likewise, the addition of a new fault or a structural feature constitutes a new geomodel. This approach preserves the integrity of the model conceptually while allowing for iterative updates and refinements without altering its fundamental identity. It is aligned with the definition of a space-time geomodel, which must satisfy a predefined conceptual model. Therefore, multiple space-time geomodels may satisfy the same conceptual framework, offering flexibility in interpretation.

Although this approach allows keeping a certain integrity of the geomodel and further refinements, one issue could be encountered is the large number of conceptual models that could be generated when adding numerous types of entities (e.g., faults or folds) during the modeling process. For further flexibility, POKIMON allows that a conceptual geomodel could satisfy another conceptual geomodel, and in a modeling context, there must be defined a general conceptual model that comprises the overall stratigraphic relations and types of major structures (as it is often considered by 3D modeling experts). This general conceptual model will be after that satisfied by the various conceptual models generated during interpretation detailing further identified geological entities and their properties (e.g., a certain number of folds in a specific folding pattern) that by their turn are satisfied by the corresponding space-time geomodels. The idea here is to move the criteria for the geomodel integrity to be applied on the general conceptual model, thus more flexibility to add/modify entities and their types while keeping a general integrity related to their general conceptual definition.

Relation with other geomodels

As discussed in the use case 3, a geomodel may be replaced by another for various reasons: to convey a different level of detail, to provide a type of information compatible with a specific representation system, or to substitute it with another geomodel that better satisfies certain constraints. To capture these aspects POKIMON employs three relations between geomodels:

- *Substitutes*: a general relation to be used between a geomodel that is proposed instead of another one as the substituted geomodel does not satisfy all information entities (e.g., contains structures smaller than the predefined size constraints) or that describe specific configurations (e.g., a duplex structure with the Geomodeller software) not supported by the deployed 3D representation system (Figure 3-14A). The substituted geomodel and the described entities are preserved in the knowledge base and can be later reused if needed.
- *Simplifies* and *Details*: These relations are a specification of the *Substitutes* relation and to be used inversely between two geomodels that describe the same geological configurations (same conceptual geomodel) and can be represented by the same geomodelling system but that are used for different purposes or at different resolutions scales (Figure 3-14B). The simplified and simplifying geomodels are related as when geomodel A *simplifies* geomodel B; geomodel B *details* geomodel A.
- *Satisfies*: a relation that indicates that the geomodel satisfies (conform to) an information entity (namely norms and constraints) or to an information content entity (namely a geomodel) (Figure 3-14A). The idea here is that a conceptual geomodel describing the structural properties of a given study area can satisfy a general geomodel describing the overall geological properties of the type of this area. For example, a geomodel describing the Peace River Arch (O'Connell et al., 1990) can be modeled in POKIMON as a geomodel that itself satisfies a global geomodel that describes the global genesis of arches areas.

The *substitutes* relation is to be used exclusively between geomodels (conceptual or space time). On the other hand, the *Simplifies* and *Details* relations are be used only between The Depiction Content Entity as they are intended to provide only a simplified/detailed depiction of the same geomodel.

Propagated and transitive properties

Various properties are propagated to geomodels from the entities associated with them. This section focuses on two key propagated properties: the localization of the geomodel and its origin. Both conceptual and space-time geomodels have the disposition to be represented using an ersatz space-time localization –a location different from the original position of the geological entities they represent. For example, a map depicting a specific location in the Pyrenees may physically reside in an office in Orleans. Similarly, a stratigraphic column could be drawn on a piece of paper and moved between offices. This ersatz location is a property that is propagated from the information artifact on which the geomodel is concretized.

Other properties, such as aboutness relations (which point to the subjects of the geomodel), satisfaction (when the geomodel adheres to norms and constraints), and parthood relations (such as the "has semantic component" relation) are also propagated. These properties follow the transitivity mechanism as when an entity A is related to B by relation R, and B is related to C by the same relation R, then A is also related to C by R. However, for computational efficiency, these transitive properties are not always explicitly modeled, as doing so can significantly increase computation time.

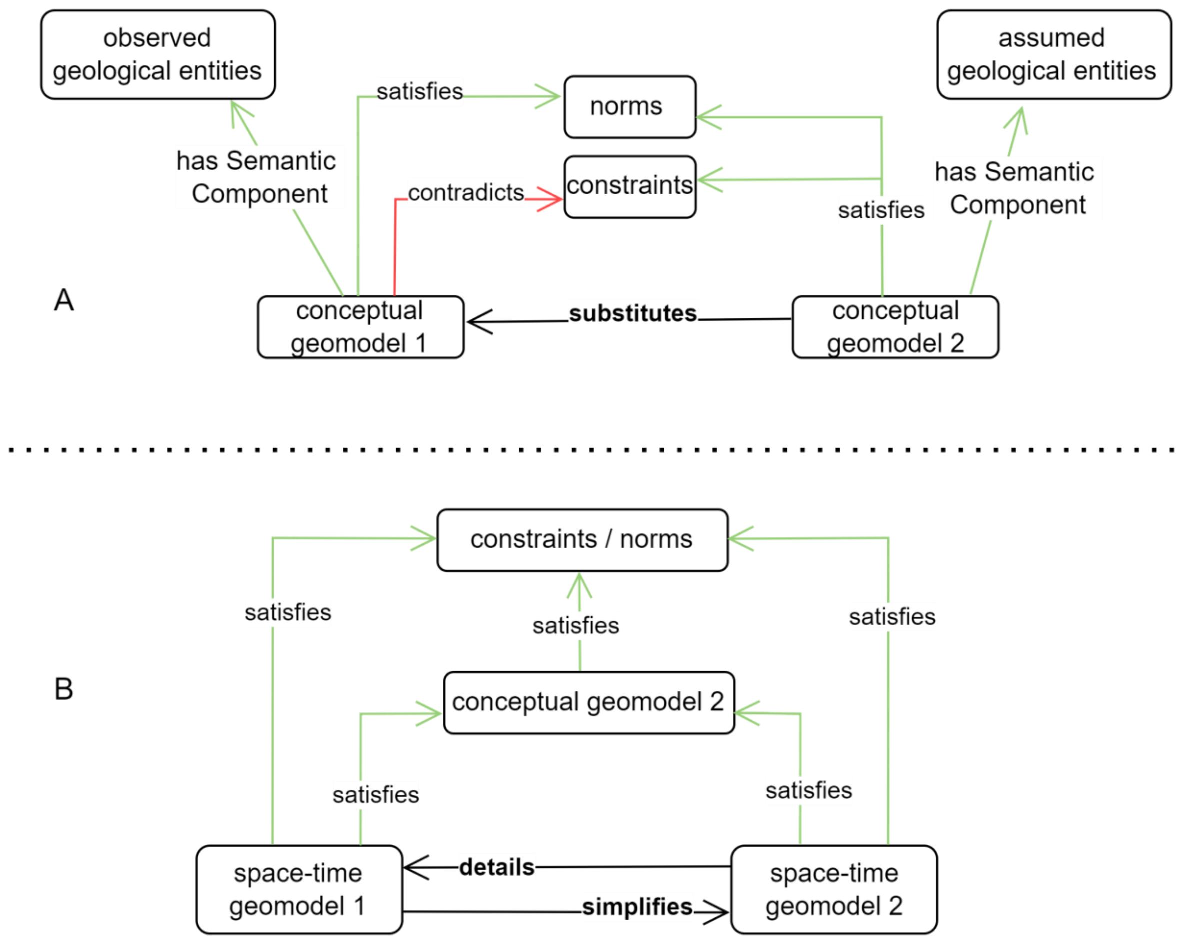


Figure 3-14 Schematic diagram to illustrate the relations of substitution and simplification between geomodels. Panel (A) displays conceptual geomodel 1 that satisfies real world geological entities, though these entities do not satisfy the constraints of the study, this geomodel is substituted by conceptual geomodel 2, which is about some assumed entities. Conceptual geomodel 1 can be stored in a local knowledge base and recovered if needed later. Panel (C) displays two space time geomodels (1 and 2) which simplifies and details one another respectively. They both satisfy the same conceptual geomodel; but different in how they describe the properties of the geological entities.

3.4.2 Processes in POKIMON

A process is an entity that occurs over time and brings about changes in the world. Processes can be either planned or unplanned. A planned process follows a plan specification that prescribes how it will unfold. For instance, when geomodelling experts decide to create a geomodel, they have a specific plan in mind that guides their actions. In contrast, unplanned processes occur without a formal plan specification, such as when a rock falls from a cliff after a rainstorm due to gravity. These processes are governed by norms, which dictate their behavior. Some norms are fully understood by current scientific knowledge (e.g., gravity), while others are known to exist, but their precise mechanisms are harder to fully grasp (e.g., radioactive decay of atoms). To model Planned Processes, POKIMON uses the IOF framework, where a plan specification (a subtype of the ICE) is linked to a process through the relation prescribes (Figure 3-15). The plan specification has two parts: (1) an action specification, which details the steps to be performed, and (2) an objective specification, which defines the desired outcome or purpose. Both are also subtypes of ICE.

POKIMON introduces further constraints on this framework. The plan specification is specifically dependent on each of its components, meaning that the identity of the plan relies on these components. Two plans with identical components are considered the same, while replacing any component results in a different plan. On the other hand, an action specification can apply to zero or one plan, meaning it is unique to each plan, whereas the objective specification can apply to multiple plans and be shared among them. For example, multiple experts working on separate geomodels for a region may share the same objective (i.e., to produce a consistent model), but their individual plans may differ in terms of data preparation, interpretation, interpolation, and representation. Furthermore, each planned process individual is prescribed by a single plan and is specifically dependent on it, but a planned process type potentially is satisfied by multiple plan specifications. An implementation therefore can query a type for appropriate plans to possibly execute, or query objectives and determine potential plans and types. In this way POKIMON allows experts to achieve the same objective by performing the same type of process in different ways.

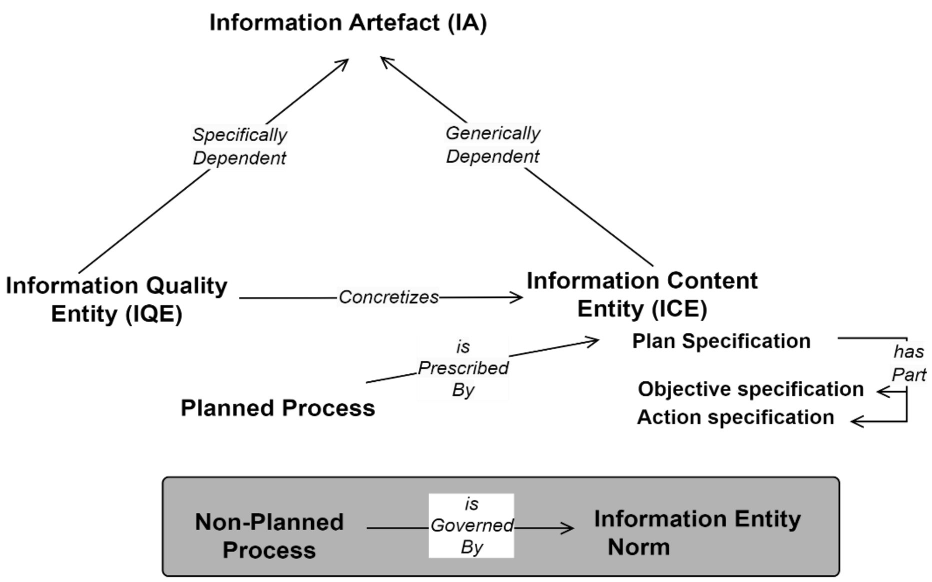


Figure 3-15 Schematic diagram depicting the modeling of the processes in POKIMON using the IOF framework. Planned and Unplanned-Processes. Planned Processes are prescribed by plan specifications subtype of ICE. Non-Planned processes are governed by norms subtype of the information entity

For unplanned processes, POKIMON links them to the norms governing their behavior through the relation is governed by (Figure 3-15). The norm itself is an information entity, as explained earlier in section 3.4.1.d. An unplanned process type or individual can be governed by multiple norms that determine how it unfolds.

3.4.2.a Participants and parts

Both planned and unplanned processes involve participants that have roles as inputs and outputs. Some processes add, entities, delete or update from existing ones.

- In an interpretation process, the input may be instances of the class Situation, and the output could be instances of the class Geomodel, which offers an explanation for the situation.
- In a selection process, initial inputs of any type are provided, with the output being an aggregate collection that may consist of some or all of the inputs.
- A folding process takes geological units as inputs and generates outputs consisting of the same units but with additional features, such as folds that occur in the input units.

Processes can also have parts that are only processes, but they may be planned or unplanned. For example, consider a person throwing a rock off a cliff. The act of throwing the rock is a planned process because it has an objective, even if the objective is trivial. This planned process is composed of several sub-processes, such as grabbing the rock, aiming, and throwing it along a trajectory. It also includes unplanned sub-processes, such as air friction affecting the rock or biological processes occurring within the person’s body. The level of detail in modeling these processes depends on the modeler's goals and the required granularity of the study.

3.4.2.b Geomodelling processes

In the context of geomodelling, all processes are currently modeled as planned processes (appendix B). The key processes related to interpretation are detailed below, based on the work of Laouici et al. (2024):

- *Selection Process*: This process selects a collection of entities. It accepts inputs of any POKIMON instances and outputs an aggregate object. The selection process is used during various stages of geomodelling, such as choosing initial inputs for interpretation or selecting a candidate explanation among multiple options.
- *Adding Instance Process*: This process models the creation of a new instance in the knowledge base. It takes no inputs but outputs a newly created POKIMON instance.
- *Omitting Instance Process*: This process models the deletion of an instance from the knowledge base. The input is the instance to be deleted, but there is no output. To preserve information, the deleted instance may be described in the annotation section.
- *Modifying Instance Process*: This process models changes made to an instance, including its properties and relationships with other instances.
- *Evaluation Process*: This process evaluates a geomodel based on its conformity to various information entities such as norms, and constraints as discussed earlier. The inputs are the geomodel and these information entities, while the output is an *Evaluation*

Entity. The evaluation entity, modeled as a subclass of the Quality class, is associated with the geomodel being evaluated. Currently, POKIMON only captures the inputs and outputs of this process, but future developments may include reasoning capabilities to model the evaluation process more comprehensively (see Discussion).

- *Decision Process*: This process models the decision-making process regarding a geomodel's fate (e.g., whether to delete or keep it). The inputs are the geomodel and its evaluation, while the output is a decision entity, which is a subclass of the Information Entity class. The decision entity is linked to the geomodel through the *is About* relation.
- *Estimation Process*: The estimation process is used when experts estimate parameters of geological entities. The process could involve simple mathematical operations (e.g., calculating averages) or more complex techniques such as interpolation. When used during interpretation, these estimation operations are modeled as processes. The idea here is to model when these operations as processes, principally interpolation when they are used to generate information during the interpretation. When these techniques are applied to generate the geometric representations of already interpreted areas as when geomodels are reproduced, this interpolation is not modeled. The estimation process takes some inputs with mathematical values and some already identified geological entities missing some values, and outputs some values attached to these entities.
- *Interpretation Process*: This process models the creation of an interpretation as described by Laouici et al. (2024). It takes a situation as input and outputs a geomodel. The interpretation process is composed of several sub-processes, including selection, adding, modifying, deleting instances, evaluation, and decision processes.
- *Representation Process*: The representation process is modeled when the geomodel needs to be expressed in different representation modes from the one in which it was initially created. The geomodel is generated through an interpretation process, resulting in a geomodel that is an ICE (Information Content Entity) generically dependent on a representation. By nature, the geomodel is inherently linked to a representation upon its creation. However, to manage the geomodel semantically, as experts do, it is first generated at a semantic level within the ontology, much like how experts form interpretations in their minds. When this ontological representation needs to be externalized through a specific representation mode –using one of the Depiction Content Entities– the representation process is modeled. It takes a geomodel as input and outputs the desired DCE.
- *Reasoning Process*: POKIMON also models the reasoning process as a planned process, allowing to capture when experts employed specific reasoning operations to infer critical information during the interpretation. For example, when experts employed an analogic reasoning to infer properties of a fold based on other folds, this process could be models with inputs being the fold with known parameters, and the folds without parameters, and the outputs being the second fold parameters. Currently, POKIMON only captures the inputs and outputs of this process, but future developments may include reasoning capabilities to directly apply some of them within the ontology (see section Discussion).

3.4.2.c Algorithms

Geomodelling processes could be run by human operators or by systems. In the case of using systems, plan specifications must be further specified, for instance in the form of algorithms. An algorithm is thus considered in POKIMON as a subclass of the type plan specification. Action specifications for algorithms in geomodelling are further distinguished into the following subtypes:

- *Deterministic specification*: defines step-by-step actions with precise instructions that must be followed in a specific order, leaving no room for variation.
- *Probabilistic specification*: includes actions that have a range of possible outcomes and are executed based on certain probabilities.
- *Conditional specification*: which outlines actions that are dependent on specific conditions or events, allowing for more flexible responses to changing circumstances.
- *Random specification*: involves actions that are chosen randomly from a set of possible actions, introducing an element of unpredictability.
- *User-based specification*: relies on inputs and decisions based on user preferences
- *Machine learning specification*: involves actions determined by models that learn and adapt from data over time with a black-box component.

Depending on how action specifications organize the different actions, the prescribed process is qualified accordingly (e.g., deterministic process, random process) using a relative quality. It is worth noting that although it is not the purpose of POKIMON to underpin simulation engines, semantic entities about geological processes might be modelled as plan specifications, and thus co-exist with geomodel in a POKIMON implementation

3.4.1 Summary of entities proposed in POKIMON

In this section, I summarized the main entities (classes and relations) proposed within POKIMON, and those adopted from existing ontologies. By adapted I mean that their range or domain are slightly updated to fit POKIMON needs (see next section for implementation and conformance with source ontologies). See the link (POKIMON) to explore all the entities implemented in the OWL file.

Classes:

- *Geomodel*: An *Information Content Entity (ICE)* that represents real geological entities. It consists of: a Semantic Component and a Depictional Component.
- *Semantic Content Entity (SCE)*: An *ICE* providing an abstract description of geological entities. It is linked to geological entities via the has Semantic Content relation.
- *Depiction Content Entity (DCE)*: An *ICE* that describes the depictional content of the geomodel.

- *Conceptual Geomodel*: Defines geological concepts and their structural relationships without full spatiotemporal details. It captures topology and relative time but lacks precise geometry and location.

- *Space-Time Geomodel*: Represents a specific geological configuration with precise spatial and temporal properties. Each Space-Time Geomodel satisfies a Conceptual Geomodel.
- *Information Entity*: An abstract *Continuant* that exists independently of representation, such as natural laws or scientific theories.
- *Norm*: A subclass of an information entity that governs behaviors, including geological laws, principles, and theories.
- *Constraint*: A subclass of an *Information Entity* defining criteria for geomodels, primarily setting spatial quality limits (e.g., length, thickness).
- *Situation*: An *Independent Continuant* that includes *SCEs* as parts. It represents a geological context where certain attributes are incomplete or deviate from norms, requiring interpretation to hypothesize or modify geological entities for coherence.
- *Origin*: A quality in POKIMON that qualifies geological entities based on their certainty level into three types: observed, interpreted.
- *Evaluation*: A quality that inheres in a geomodel, representing the assessment outcome of its validity. It can have qualitative (e.g., good/bad) or quantitative (e.g., probability) values.
- *Decision*: A subclass of realizable entity that determines the fate of a geomodel (e.g., accepted or rejected) based on evaluation results. It is specifically dependent on and about the evaluated geomodel.Algorithm
- *Unplanned Process*: A process that occurs without a formal plan, governed by norms.

Relations:

- *Substitutes*: A relation where a geomodel replaces another due to limitations in resolution or representation.
- *Simplifies/Details*: Inverse relations between geomodels at different resolution scales for varying purposes.
- *Satisfies*: A relation where a geomodel conforms to an information entity or a another geomodel.
- *Contradicts*: a relation where a geomodel is not conform to an information entity or a broader geomodel; it is also used between a space time geomodel not conform to a conceptual geomodel.
- *has Semantic Component*: Links a geomodel to its *SCE* and connects the *SCE* to entities in the geological domain ontology.
- *has Depiction Component*: Links a geomodel to its *DCE*, and the DCE to various representational formats.
- *Is About* : an originally BFO entity used to conceptually links a model to real geological entities, though it remains abstract and non-implementable. Its domain is updated to include the new added class Information Entity.

3.5.Implementation, application, evaluation

3.5.1 Implementation

POKIMON is implemented in the ontology web language OWL (Antoniou and Harmelen, 2009; Breitman et al., 2007) using Protégé (Noy et al., 2003), and can be found at this link (POKIMON). Although POKIMON uses entities definitions from BFO, IAO, and IOF and follows generally a similar structure, extracted entities were not directly copied with their original URIs, instead, Instead, entities are replicated with modified URIs. This approach is adopted for two main reasons. First, although POKIMON aligns with the foundational definitions provided by BFO and IAO, some definitions remain open to discussion and potential refinement. We believe that certain adjustments, even minor ones, could improve clarity and applicability—not only for POKIMON but also within the source ontologies themselves. Notably, BFO has undergone multiple revisions to reach its current structure, reflecting an ongoing process of refinement. Our updates do not introduce major conceptual changes but rather minor adjustments. However, directly copying all axioms from these ontologies, and updating some of them could lead to logical inconsistencies during reasoning. Second, some axioms in the source ontologies are not particularly useful for our case and including them would add unnecessary complexity for POKIMON users. Conversely, omitting these axioms can also introduce logical inconsistencies in reasoning. Thus, our approach carefully balances the need for alignment with foundational ontologies while ensuring POKIMON remains both practical and logically consistent.

It is important to note that, for engineering purposes, the satisfaction relation between the semantic component entities (instances of the geology ontology) and the type of instances it must satisfy is not implemented in this version of POKIMON. This is because OWL-DL constraints the instantiation relation between a type and its instances and introducing any other relation would violate these constraints. In this way, POKIMON remains logically consistent and compatible with various reasoners, such as Hermit (Glimm et al., 2014) and Pellet (Sirin et al., 2007). To demonstrate POKIMON's functionality, the next section presents two figures showing applications the use cases (1 and 4). The figures of application on the remaining uses cases are provided in appendix B, with SPARQL queries to demonstrate information retrieval capabilities if POKIMON. Based on meeting the ontological requirements, POKIMON is evaluated, and possible limitations are identified for further enhancement.

3.5.2 Application

The Figure 3-16 represents the main entities employed to model the knowledge aspects evoked in the use-case 1, which consists of: a conceptual and a space time geomodels, and their representations. As illustrated in the Figure 3-16, the *space_time_geomodel_folded_surfaces* is satisfying the *conceptual_geomodel_folded_surfaces*. Both geomodels are about the same geological entities (folded boundary1, 2, and 3), though for figure clarity only relations to the space time geomodel are displayed. Both geomodels have representations: the representation of the conceptual geomodel is provided by a cartoonish representation, instance of the Figurative DCE, while the representation for the space time geomodel is provided by two DCE instances, a textual (proposition) and a 3D figurative representation.

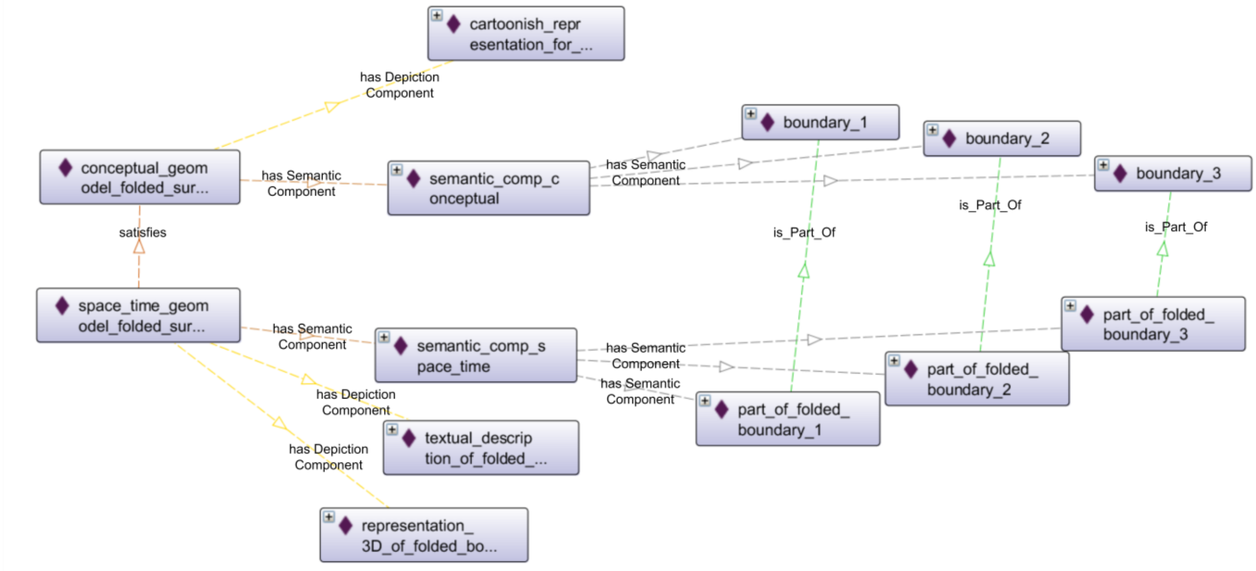


Figure 3-16 Snapshot from Protégé software showing the ontological representation of two geomodels: a conceptual geomodel and a space-time geomodel. Classes are shown with yellow patches, and individuals with purple patches.

The Figure 3-17 presents an interpretation process explained in use-case 4. Only key entities directly related to this process are displayed for figure clarity. This interpretation process (process_interpretation_fold_1) takes inputs a series of geomodels associated with initial observations (gmd_1 to gmd_10), and outputs a geomodel_fold1. This resulting geomodel is about the geological boundary 1 visualized by the yellow surface as in Figure 3-7 that was proposed, evaluated and validated within this interpretation iteration. The process_interpretation_fold_1 *is prescribed by* an algorithm of interpretation. Five subprocesses are parts of this general interpretation process, describing the main interpretative tasks as per Laouici et al. (2024).

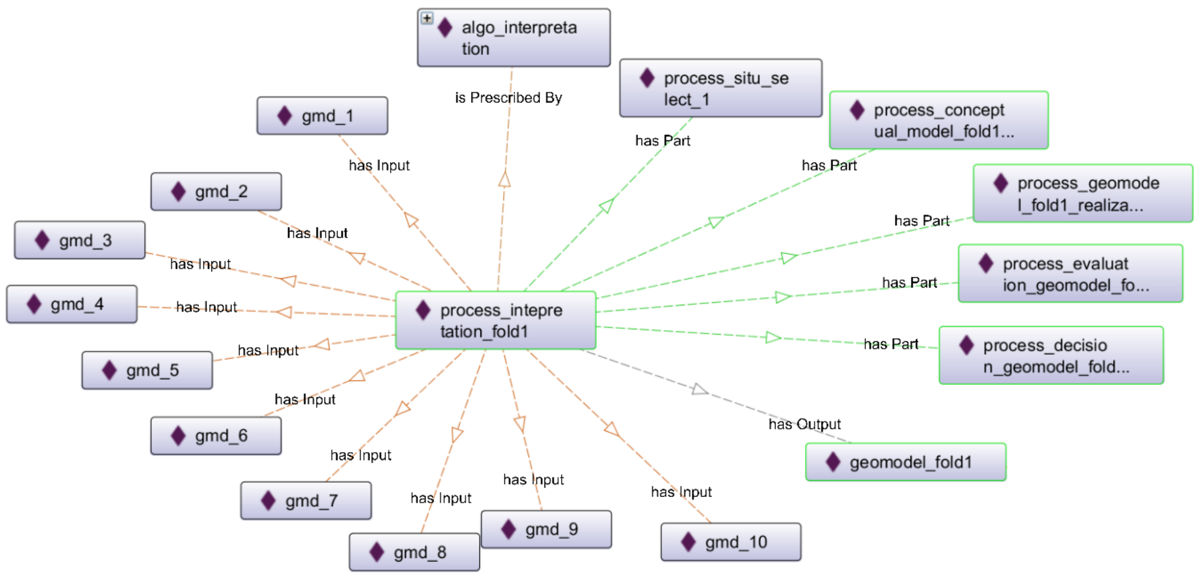


Figure 3-17 Snapshot from Protégé software showing the ontological representation of an interpretation process (process_intepretation_fold_1). Classes are shown with yellow patches, and individuals with purple patches.

3.5.3 Evaluation

The presented examples (figures and SPARQL queries) demonstrate how POKIMON answers the ontological requirements derived from the use cases (see Section 3.2 use cases). The requirements I to III pertain to the modeling of the different considerations and relations of the geological model at the conceptual level and when it is realized and represented with its full spatial properties. These requirements are answered through the ontological definitions of key geological entities (*Semantic and Depiciton Content Entities*, *Conceptual Geomodel* and *Space Time Geomodel*) as shown in figures 3-18, 3-19, 3-20 using an adapted framework from existing top and mid-level ontologies (BFO and IAO). The requirements IV to VII focus further on characteristics of the geomodels and their mutuel relations when they are updated, modified or entirely replaced during the interpretation process. These requirements are answered by defining a framework to consider the model integrity and the possible relations (*details and simplifies, satisfies*) (Figure 3-20). In addition, the origin of the information in the geomodels is modeled (*observed, interpreted*, and *assumed*) to further characterize these information.

The requirements VIII, from to XI, and XIII concentrate on the modeled geological real-world entities and they are answered by merging concepts from domain geosciences ontology (GSO) and providing framework to model the knowledge about them how this knowledge must be satisfied using specialized ontological entities (*norms, constraints, satisfies, contradicts*). Also, POKIMON provides the framework to model situations, requiring interpretation, emerging from the knowledge about geological entities. Finally, the requirements IX, X, and XIV regarding the different processes and their specifications and how they interact with different entities in the modeling framework are answered by providing a common framework to model them using key concepts from the IOF ontology that were extended to cover the algorithmic aspect (Figure 3-20).

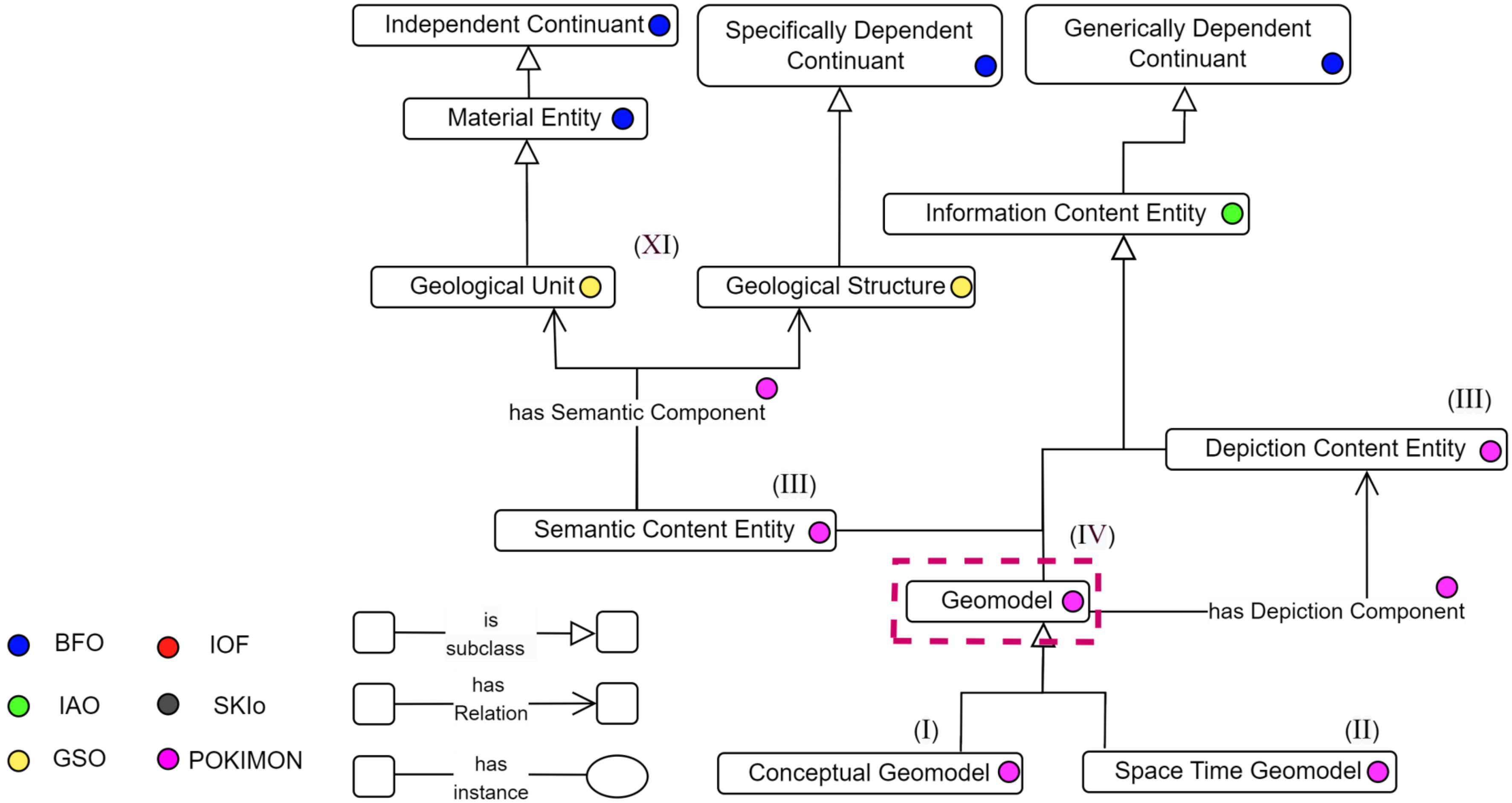


Figure 3-18 Schematic plate depicting the main concepts of POKIMON defined by the ontological requirements stated above (noted by the Roman numerals)

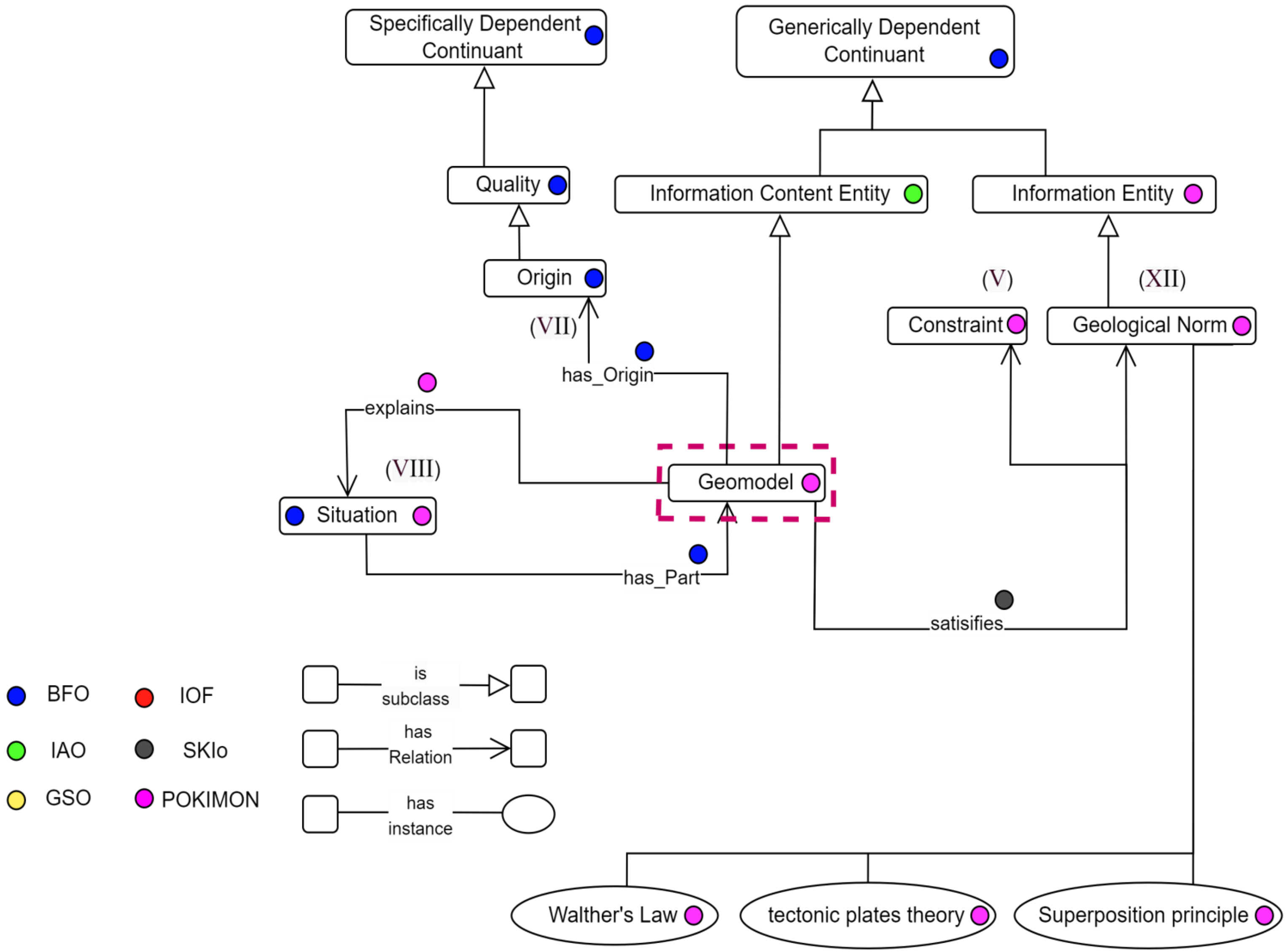


Figure 3-19 Schematic plate depicting the main concepts of POKIMON defined by the ontological requirements stated above (noted by the Roman numerals)

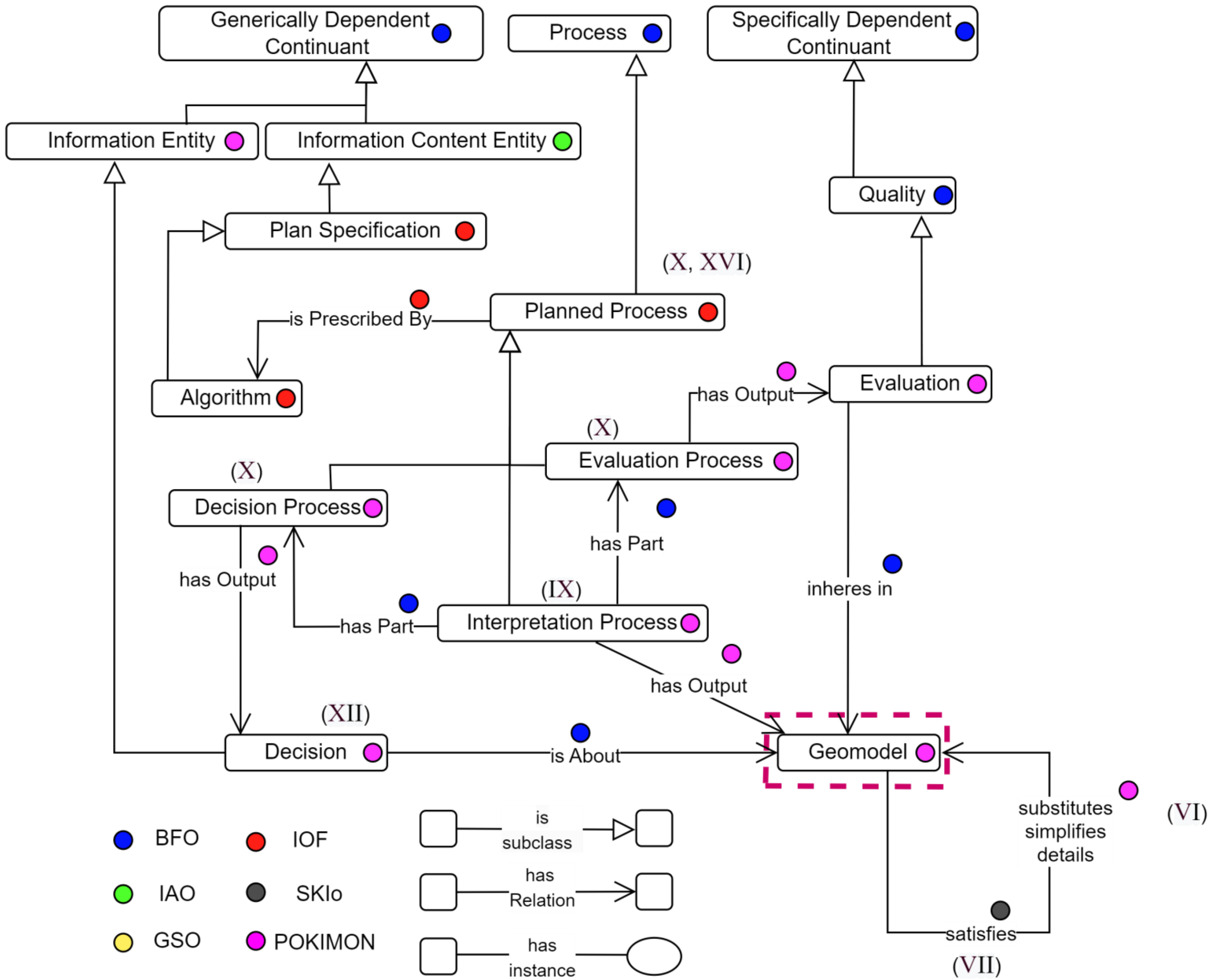


Figure 3-20 Schematic plate depicting the main concepts of POKIMON defined by the ontological requirements stated above (noted by the Roman numerals)

3.6. Discussion

3.6.1 POKIMON achievements

As illustrated by the application results on the use cases, POKIMON offers a flexible and generic approach to capture different facets of expert's knowledge during the geomodel construction. This knowledge is employed and generated by human operators as often done with classical geomodelling systems where experts are in charge of the interpretation process or when systems contribute firmly to the interpretation process as per Laouici et al. (2024). POKIMON brings a generic and consistent framework to define different types of geomodels based on their structural characteristics and the representational properties either on a conceptual or realization level.

In the 3D geomodelling community, geomodels are often defined as the combination of inputs, included concepts and interpolation algorithms (Calcagno et al., 2008; Loiselet et al., 2016). While such definition is pragmatic and allows adequate storing and querying of geomodels, it poses an ontological challenge. One key challenge is that geomodels, when described only by their inputs and algorithms, are reduced to a functional perspective without fully accounting for their inherent ontological nature. This perspective overlooks the deeper conceptual distinctions between what a geomodel is, how it is created and what it does. For instance, the same semantic network described by a geomodel could be generated by multiple systems using different estimation algorithms or interpreted by two different experts. Moreover, considering only functional properties differs from one expert to another as a geomodel could be seen as a representation of geological knowledge, a computational tool, or a means for simulation. Each of these perspectives impacts how the model should be classified, stored, and queried. Without a rigorous ontological framework like POKIMON, there is a risk of conflating these roles, making it difficult to differentiate between the conceptual model and its realizations or to support consistency across various applications. POKIMON acknowledges the need for storing and describing the associated information associated with a geomodel (inputs, and algorithms), and structures them according to their ontological relations with geomodels. In this way, POKIMON provides a flexible framework for creating, storing, and querying geomodels and associated knowledge.

By providing a framework to record the origin (observed, interpreted, assumed) of each information and model part and making links between the different geomodels developed to substitute, simplify or replace other geomodel, POKIMON allows tracing the entire history of the construction of the geomodel and the origin of information which can significantly help during the interpretation process. Such challenges are significant in the field of 3D geomodelling and often poses issues when multiple experts build models collaboratively, and often undertaken from an engineering perspective (Lidal et al., 2013). In addition to addressing these domain challenges, by doing this, POKIMON addresses specifically the challenges posed by Whitehead and Gahegan (2012) about the epistemic and uncertain aspects of the geoscientific information that are often not addressed by ontologies in the field of geology.

POKIMON is an independent OWL encoding but that achieves formality using existing well-founded ontologies such as BFO (Otte et al., 2022), IAO (Ceusters and Smith, 2015), IOF (Drobnjakovic et al., 2022), and GSO (Brodaric and Richard, 2020). POKIMON is positioned as a bridge between ontologies that describe geology at a general level and those that describe information and its representation. POKIMON integrates them in a comprehensive and coherent framework to ensure that the interpretive processes and algorithmic procedures involved in the creation and operation of 3D geomodelling systems are thoroughly encapsulated. This integrative approach supports the formal representation of the knowledge and logic underlying geomodelling tasks.

3.6.1 POKIMON and conceptual modeling foundations

This part of the work extends foundational contributions of Perrin and Rainaud (2013), who explored the definition of various notions related to geological models (models, representations, and visualizations). It also builds upon Abel et al. (2019), who formalized knowledge for categorizing visual content in petroleum engineering. These works of Perrin (2013) and Abel (2019), align with the perspectives of Guizzardi (2005) and Guarino et al. (2019), who consider models as mental constructs that are later externalized through model specifications using a modeling language. In this view, a model (the mental model in Guarino's framework) is an organized instance of a conceptualization (set of concepts), which is then represented as a model specification via a modeling language.

POKIMON, however, adopts a different ontological stance (Figure 3-21). Instead of treating a model as a conceptualization, later represented, POKIMON follows the IAO framework by collapsing the conceptualization and the model into a single subclass of an *ICE*—specifically, a *Geomodel* that has two component, an abstract one the *Semantic Content Entity,* and a tangible one *a Depiction Content Entity*. In this framework, a model inherently requires representation to exist. Consequently, the same model may exist as a mental representation within multiple geologists' minds, even if they employ different conceptualizations to reach it. The modeling language described by Guizzardi (2005) and Guarino et al. (2019) aligns with POKIMON's *Depiction Content Entity*, which serves as the externalized component of a model. Additionally, following IAO classification, the model specification in Guzzardi's (2005) is delineated into the *Information Artifact* and the *Information Quality Entity* that allow concretizing the model. By adopting this approach, POKIMON allows preserving a common understating of the notion of a conceptual model within the geological field—a simplified description of a geological configuration (Chambefort et al., 2016; Heather R. Miller and Herbert, 2010; Silva-Fragoso et al., 2021; Wolf, 1976). This conceptual model may exist as an internal mental representation or be externalized in a simplified, cartoon-like form.

POKIMON merges two complementary ontological viewpoints on modeling. First, it adapts the foundations laid by Guizzardi (2005) and Guarino et al. (2020) to position POKIMON itself as both a model specification (Guizzardi's perspective) and a conceptual model (Guarino's perspective), formalized within an ontological framework. Second, POKIMON explicitly integrates the inherent representational component of models, enabling it to address domain-specific application requirements –geomodels– while providing a hierarchical structure for their formalization and classification

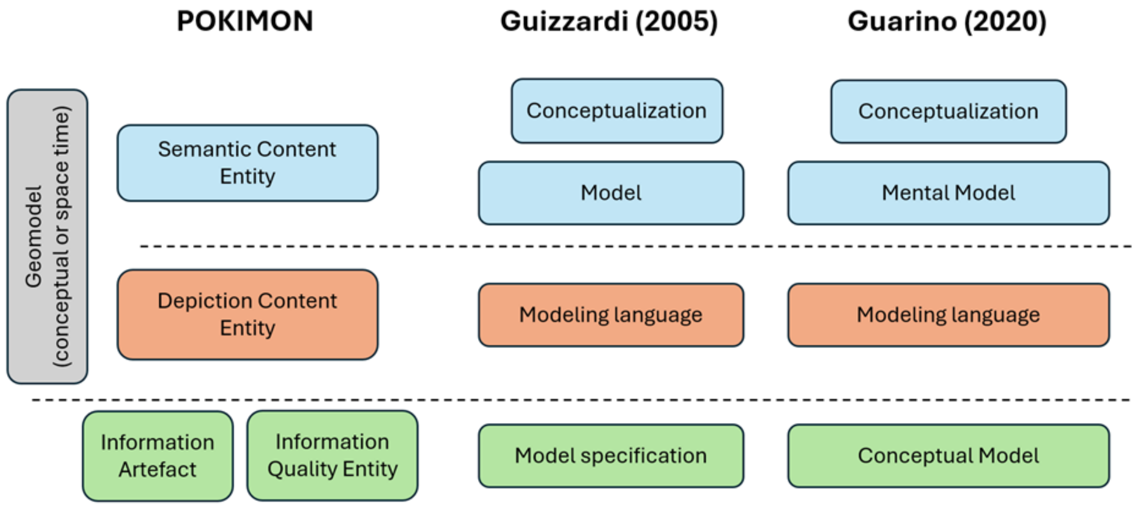


Figure 3-21 comparison of the proposed definition of geomodels with the main entities used in conceptual modeling in Guizzardi, 2005 and Guarino et al., 2020 (see text for detail)

3.6.1 Important considerations

An important consideration is that geomodelling experts, while applying their understanding of types of geological entities and adhering to established geological norms, do not discover new types of entities or establish new norms. Instead, they identify potential instances of geological entities governed by these pre-existing norms. From an ontological perspective, this means that interpreters (humans or systems) will not create new classes or define new types of relations between these classes. Instead, they will operate on a common sealed knowledge model that describes the geological phenomena.

Accordingly, in a geomodelling system, geological norms and definitions of types represent a constant part of the knowledge model also known as the TBOX, and the built geomodel corresponds to the populated knowledge base structured by the knowledge model in question also known as the ABOX (Rudolph, 2011) (see Section 1.3.1). The idea behind the development of the POKIMON ontology is to model how this knowledge base is populated regardless of the knowledge model used to describe geological entities. At this early stage of development to demonstrate feasibility, the description of geological entities in POKIMON is an extraction of a set of GSO entities. However, the framework of POKIMON allows it to integrate semantic description of geological entities from other ontologies, such as the GeoCore ontology (Garcia et al., 2020) without significant change in the critical structure of POKIMON. For example, a notion to be considered if POKIMON used GeoCore would be the difference in classifying geological structures. POKIMON with its focus on the interpretation aspects, consider them as specifically dependent continuants allowing more localized and entity-specific interpretations, while GeoCore considers them as generically dependent continuants emphasizing the generalizability and repeatability of geological structures, allowing them to be understood in a more generic sense which goes along with the nature of the GeoCore ontology.

3.6.2 Limitation and challenges

One major limitation of POKIMON is its incompleteness in various areas. For instance, the interpretation process employed by experts is a complex and tangled process that has not yet been properly formalized in a conventional algorithmic way within the geomodelling community or applied in modelling tools, and depends on software implementation. Work such as that by Laouici et al. (2024) addresses this issue through the structure of this process and the use of explicit knowledge. The proposal in this chapter provides a framework that would help in further formalizing this process explicitly. Different geomodelling workflows and tools may use various processes. Although POKIMON offers a general and flexible framework to model these processes, there is a need for a more exhaustive listing and specification of them by domain experts. To achieve this, comprehensive surveys on geomodelling tools, practices, and user preferences need to be established in the future.

Another limitation of POKIMON is the general weaknesses that could be associated with its top-level ontologies, particularly the classification by BFO of continuant entities as generically dependent, specifically dependent, and independent entities that are mutually disjoint. The issue is that this framework constrains modeling entities that could be generically dependent on other entities while being specifically dependent on others. One solution is to basically avoid such hierarchical classification and use only the generic or specific dependencies as object properties allowing further flexibility as employed in GSO (Brodaric and Richard, 2020).

3.6.3 Extending POKIMON for reasoning

This section introduces preliminary insights into utilizing POKIMON for reasoning and its collaborative application during the geomodelling process. A considerable amount of discussion and reflection has been dedicated to this topic, and I believe it merits highlighting within this section. The POKIMON ontology framework enables the ontological representation of various processes involved in geomodelling, especially those associated with interpretative tasks, including the mental reasoning processes employed by experts to deduce properties or abduce concepts. However, POKIMON does not yet contribute to the reasoning processes needed to infer qualities of modeled geological entities, propose candidate explanations for observed situations, or identify why certain geomodels do not satisfy their associated information entities (e.g., geological norms). Currently, POKIMON is primarily used to check the consistency of the populated knowledge base and infer various types of information, such as inverse relations or the classification of individuals based on asserted axioms.

Reasoning with POKIMON can be further developed to infer critical information that supports interpretative tasks and partially automates the reasoning processes typically employed by experts. Such development on reasoning capabilities can be achieved by focusing on two key areas: (1) reasoning to infer possible theories that explain observed situations, contributing to the interpretation process, and (2) reasoning to infer why geomodels fail to satisfy information entities, contributing to the evaluation process. These two reasoning areas are interrelated and require an extension of POKIMON's current structure.

3.6.3.a Reasoning to Propose Theories

As noted by Laouici et al. (2024), a potential explanation relation (*is Possible Explanation Of*) can be established between different geological entities and their potential explanations, enabling the reasoning engine to infer possible candidate explanations based on this relation. This is distinct from the *explains* relation used in POKIMON, which exists between a specific unexplained situation and an already asserted explaining entity. In contrast, the *is Possible Explanation Of* relation is applied between a situation and all possible explanatory entities. From an ontological perspective, as discussed in this chapter, the *explains* relation exists only at the modeling level –between geomodels– and not between real-world geological entities that possess natural relations. This distinction is crucial because experts interpret not only situations arising from the natural behavior of geological entities but also situations resulting from erroneous prior interpretations or inaccurate estimations caused by incomplete or low-resolution inputs. Thus, what experts interpret is a mixture between situations requiring proposition of explaining entities, and others requiring modifying existing entities. Accordingly, establishing possible explanation relations requires undertaking all these situations and modeling them based on their geological and geomodelling characteristics.

The work of Parquer et al. (2024) parallels this endeavor by focusing on detecting inconsistencies (or, using the vocabulary of the current research, inconsistent situations that arise during modeling), but with no consideration of their ontological implications. This initiative considers geological principles and entangles spatiotemporal norms that structural models must satisfy to be considered valid.

Therefore, modeling situations from an ontological perspective in POKIMON to perform reasoning can systematically benefit from these foundations (Parquer et al., 2024) and ontologically enumerating these norms. It also requires further interest in the origin of information behind the situation to distinguish between situations that can be explained by additional entities to relate them to their possible explanations as they are acceptable at the geological level, and those that are geologically invalid and need to be omitted and relate them to pertinent processes.

3.6.3.b Reasoning in the evaluation process

As one may perceive, establishing the possible explanation relations to situations has direct impact on the evaluation process as it already establishes specific paths to verify the conformance of these situations. Using reasoning in conjunction of these situations and their possible explanations enables during evaluation capturing whether a situation satisfies or contradicts them and the geological reason behind. This raises the reasoning process to a meta-level, where the reasoner does not halt due to errors but instead generates possible relations and reclassifies anomalous situations, allowing the geomodelling process to continue even in the presence of anomalies. Rather than breaking when inconsistencies are detected, the system adapts dynamically, contributing to the automation of the evaluation process.

POKIMON already provides the infrastructure for structuring these aspects (proposing theories and the evaluation process), and the idea is to extend it to encompass detailed descriptions of geological norms and situations and use them within an 'Ephemeral' reasoning framework. This framework shall be applied each time an unexplained situation arises and cleared afterward, allowing it to run anew without any prior instances of situations. By doing this, reasoning focuses on a small subset of entities rather than the whole set of instances of the entire geological model, which often includes a large number of entities, most of which have already been validated in previous iterations. This approach helps optimize computational time and memory requirements, which are often challenging when reasoning with large knowledge bases and it follows the main ideas guiding the formalization of the proposed geomodelling paradigm based on progressive construction of the geomodels.

3.7.Conclusion

This chapter presents the work on POKIMON, a comprehensive ontology designed to capture the knowledge aspects employed during the 3D geological structural modeling process, with a particular emphasis on the interpretation process. By integrating foundational concepts from established ontologies such as the Basic Formal Ontology (BFO), the Information Artifact Ontology (IAO), and the Industrial Ontologies Foundry (IOF), POKIMON provides a robust structural foundation. Additionally, it incorporates concepts from the GeoScience Ontology (GSO) to accurately describe the geological entities involved in the modeling process. POKIMON facilitates understanding of geomodels, their representations, and the processes involved in building them during the interpretation phase.

This ontology not only standardizes the representation of geological knowledge but also enhances the interoperability and reusability of data across the three disciplines (geology, modeling, and knowledge representation). Through detailed examples and use cases, the efficacy of POKIMON is demonstrated, showcasing its ability to encapsulate complex geological interpretations and streamline the modeling workflow. These practical applications highlight the ontology's potential to improve the accuracy and efficiency of geological modeling, making it a valuable tool for both academic research and industry practice.

General discussion

The proposed research addresses multiple scientific challenges within two primary aspects of the 3D geomodelling field: (1) the interpretation of geological architectures and (2) the incorporation of expert knowledge used during interpretation within the geomodelling systems. These two aspects are deeply interconnected, meaning that addressing one inherently involves addressing the other. We advocate for employing conceptual modeling and formalization to tackle these challenges. While these approaches are not entirely new in geosciences, they have recently gained increased attention, especially for facilitating the sharing of geoscientific data and standardize knowledge (Abel et al., 2012b; Simons et al., 2006; Whitehead and Gahegan, 2012). Ontologies, for instance, provide robust frameworks due to their ability to formalize semantic relationships, ensure consistency in knowledge representation, and support interoperability. These characteristics make them particularly effective for storing and sharing the semantics associated with geometric properties of already constructed geomodels (Qu et al., 2023; Zhan et al., 2022).

As detailed in the State of the Science Chapter, a significant challenge remains: how to enable the dynamic use of formal knowledge by human experts during the iterative construction of geomodels? Such ambition advocates an infrastructure that supports the mutual understanding, application, and storage of information related to geological phenomena, algorithms, workflows, and interpretive decisions. This Ph.D. thesis situates itself at the intersection of these areas, aiming to bridge the gap between interpretation and the formalization of expert knowledge.

Innovations of the approach

A primary contribution of this work is the proposition of a generic interpretation formalism to structure and automate the interpretation process as a distinct entity. Such entity is formalized through four major stages 'Plan, Do, Check, and Act' that group the various interpretative tasks within a continuous improvement cycle, applied to explain some situations that are ambiguous within the modeling context. This formalism of interpretation is based on leveraging an external knowledge model that contains the geological knowledge needed to identify, describe and validate resulting interpretations, leading to semantically rich models.

As discussed in the second chapter, modeling interpretation as a distinct entity represents an innovative approach. Existing literature primarily focuses on solutions that assist experts in reflecting their interpretations through powerful representation techniques, where experts take the lead in proposing geological entities and their relationships –without record in the system– in the proposition of existing geological entities and their relations (Brandel et al., 2005; Grose et al., 2019; Laurent et al., 2016; Perrin et al., 2005; Wellmann and Caumon, 2018). In addition to automating many interpretative tasks, the proposed formalism allows recording the history of the entire process and the information used or generated, which means further encapsulation of semantics in the final geomodels. As they are recorded explicitly, these interpretive tasks and decisions can be further queried and discussed, thereby providing a framework for addressing the epistemic uncertainty inherent in the interpreted information (Bond et al., 2007).

The endeavor presented in the second chapter dedicated to structuring interpretative tasks, provides a preliminary framework to structure the essential knowledge of 3D structural modeling around the interpretation process. It highlights the need for three key aspects of knowledge: geological, interpretive, and representational. These categories should be explicitly modeled as distinct entities, preserving their individual semantics while enabling their integration within a formal knowledge representation system, such as an ontology. By leveraging ontologies and maintaining a clear separation of these knowledge dimensions, the system enables the manipulation of geological entities as instances that may lack complete full geometric properties, mimicking the way domain experts manipulate geological concepts cognitively. Furthermore, an ontological approach to modeling processes facilitates the explicit specification of different process types and their execution modalities. This enhances the flexibility and semantic interoperability of the system, improving its generic applicability compared to conventional geomodeling workflows (Calcagno et al., 2006; De Kemp et al., 2017; Fernandez et al., 2009; Mallet et al., 1989; Maxelon et al., 2009; McInerney et al., 2005).

To model the identified knowledge aspects, we propose a minimal ontology for geological interpretation (MOGI), as a feasibility demonstration. However, further refinement is needed to clarify overlaps, particularly the distinction between spatial and representational properties, using specialized information ontologies. While well understood by experts (Perrin and Rainaud, 2013), geomodelling system developers often approach these aspects pragmatically (Calcagno et al., 2008; Loiselet et al., 2016).

Building on this foundation, the third chapter of this research introduces an ontological framework to formalize expert knowledge implicitly applied in geomodel construction. The POKIMON ontology was developed to characterize geomodels, their representational properties, and the relations between different models selected based on study requirements or system constraints. Additionally, it provides a generic framework for modeling the processes involved in geomodel development, integrating specifications within a logically consistent structure. To achieve this, POKIMON extends concepts from existing top- and mid-level ontologies (Ceusters and Smith, 2015; Karray et al., 2021; Otte et al., 2022) that describe abstract entities in scientific domains, as well as the information and its representational framework.

POKIMON does not model geological entities directly from scratch but instead relies on existing geological domain ontologies (Brodaric and Richard, 2020) for further interoperability. This approach makes POKIMON unique, as it uses existing geological knowledge models (Brodaric and Richard, 2020; Garcia et al., 2020) to provide consistent geological framework, and modeling how these knowledge models are populated within the geomodelling process. Thus, it successfully provides a framework for conciliating two distinct knowledge domains: geology and modeling. Ultimately, the proposed approach makes the geomodelling process more interactive with the geocognitive abilities of experts, allowing systems to internalize and benefit from these cognitive processes through a structured, flexible interpretation framework, and the explicit representation of a significant portion of expert knowledge. This represents a meaningful step toward automating the interpretation process.

Impact within the field

Through the development and implementation of these concepts, the role of the proposed approach in 3D geomodelling is becoming increasingly clear. As illustrated in Figure 3, this conceptual modeling framework can be effectively merged with existing geomodelling systems to create a more interactive and semantically driven environment. The idea is using POKIMON to store geomodels at the semantic level. These geomodels can either be interpreted directly by experts as traditionally done or following the proposed continuous improvement formalism (PDCA). These semantically stored geomodels can be queried and converted into different formats, such as generating text descriptions with full semantic content, or as interpolation constraints such as point clouds, or meshes, based on the geometric definitions of geological entities. A specialized converter is responsible for transforming complex geological information into formats compatible with the deployed representation formalism, simplifying or replacing entities when necessary.

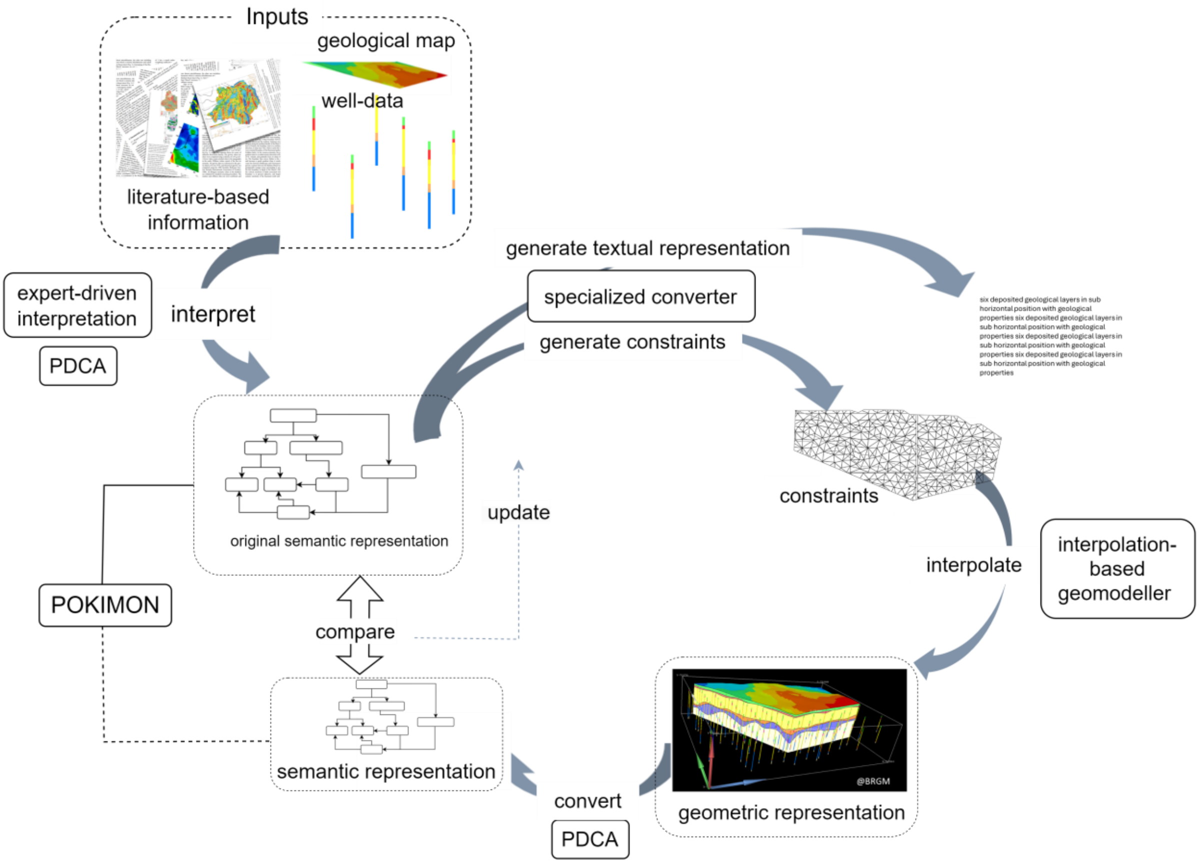


Figure 3 Schematic diagram showing the position of the proposed approach within the 3D modeling field.

The geomodelling tool can then perform interpolations to construct the geometries, which are further validated against the knowledge base and updated if necessary. This environment would not only support automated interpretative tasks but also facilitate the handling of existing geomodels, interpreting the semantics behind them for both conceptual and rendering purposes. By integrating this framework, geomodels and the knowledge embedded within them will play their inevitable role in the geomodelling process, enabling multi-representation across various rendering systems that are interoperable across platforms.

3. Challenges and general limitations

As evidenced by the second and third chapters of this research, and supported by the applications, this focus on conceptual modeling allows in fact integrating a considerable part of expert's knowledge used often tacitly and significantly advances what can be expected from geomodelling systems when they use knowledge explicitly. However, as the proposed approach is multidisciplinary it requires the incorporation of various fields, such as knowledge formalization, cognition, and epistemology, which have rarely been considered in 3D geomodelling systems outside of a mathematical framework (Wellmann and Caumon, 2018). These research areas currently lack a reusable infrastructure in the geoscientific field, necessitating substantial development from scratch. A significant portion of this research was accordingly devoted to developing the prototype algorithm and the POKIMON ontology.

To fully leverage the proposed formalism, there is room for further development regarding the decision process during the interpretation and further conceptual modeling of interpretation situations for more intelligent results. Additionally, this research can be expanded to better capture modelers' experiences and expertise to refine the strategic aspects during the geomodel construction. The work on these aspects, principally, modeling situations, formalizing them, and addressing the retroactive and prospective aspects of the interpretation process, has already begun within the timelapse of the current research and will continue moving forward.

A major challenge of the proposed approach is its deployment in geomodelling systems. Existing geomodelling systems still rely heavily on experts with no conventional way to encode tacit experts' knowledge about the geomodelling process. In addition to requiring the goodwill of software developers, particularly those in industry, to overcome resistance to change, there is a significant cultural aspect to consider. When experts develop geomodels, they often do not see the need to store their tacit knowledge because much of the process is intuitive for them. However, the issue of knowledge incorporation arises downstream in the geomodelling chain, particularly when the geomodel is used for other purposes or by other users.

While POKIMON offers a general framework –not restricted to using the proposed PDCA interpretation formalism– for capturing knowledge about geomodels and their relation to interpretation, its current implementation requires significant manual input, which could hinder widespread adoption. Automating the knowledge record process would significantly enhance usability, making it more appealing for experts to use the proposed framework as part of their workflow.

From an engineering perspective concerning the deployment of ontologies, notable challenges associated with the proposed approach include the inherent limitations of these formalisms –ontologies. The balance between expressivity and simplicity is critical, as discussed in the first Chapter. Enhanced expressivity may lead to increased computational demands and complexity in reasoning, thus complicating the maintenance and debugging of ontologies. This is why many application ontologies are often lightweight and modularized (d'Aquin, 2012; Davies, 2010; Giunchiglia and Zaihrayeu, 2007). The POKIMON ontology serves as a proof of concept for the proposed perspective on integrating and storing knowledge throughout the geomodelling process, necessitating a commitment to a certain level of expressivity. Future applications and extensions could benefit from a more lightweight variant of POKIMON to facilitate smoother integration within efficient computational operations.

Another significant challenge arises from the limitations associated with OWL-DL, particularly the inability to further manipulate instantiation relations and their satisfiability. In POKIMON, these needs are currently delegated to the user or an external system to avoid potential bugs. Furthermore, OWL-DL is constrained in its capacity to represent dynamic systems. Although interpretation is inherently a dynamic process where new outcomes are continuously generated, main issues such as reclassification have not yet been critically encountered. Therefore, POKIMON has thus far succeeded in effectively modeling such processes.

Conclusion

In conclusion, this research work proposes original contributions to the field of 3D geomodelling: (1) the interpretation of geological phenomena during geomodelling and (2) the incorporation of expert knowledge within the geomodelling systems. Traditional approaches to 3D structural modeling often treat it as an interpolation problem, requiring geological knowledge to be mathematically formulated for interpolation algorithms. In contrast, the proposed approach shifts focus toward conceptual modeling and structured interpretation. Accordingly, I develop in this research work a novel conceptual modeling framework grounded in a structured interpretation process and the formalization of expert knowledge, offering a new path to overcome the limitations of current geomodelling systems –particularly the challenge of incorporating tacit knowledge. By shifting the emphasis from purely computational geometric modeling to a more interactive, semantically-driven system, this research enables knowledge to be captured, shared, and queried at a conceptual level.

A primary contribution of this work is the introduction of a generic interpretation formalism based on continuous improvement (PDCA), which structures the interpretative process as an independent entity. This formalism enhances the geomodelling workflow by recording the history of interpretations, facilitating transparency, and fostering automation. The research also introduces POKIMON, a formal ontology designed to capture and structure expert knowledge related to geomodelling, particularly the different types of geomodels and their representations. By providing a framework to separately describe geological, interpretative, and representational knowledge aspects, POKIMON enables flexible and semantically rich geomodelling processes. Accordingly, the proposed approach holds significant potential for transforming how expert knowledge is formalized, stored, and utilized in systems. While these contributions represent considerable advancements, several areas are open for further development. Future work should focus on the refinement of strategic decision-making during geomodel construction with the PDCA. Furthermore, fully integrating POKIMON within a system that automates the knowledge recording process could significantly enhance its usability and streamline integration into existing geomodelling systems.

This research paves foundations for future developments in the 3D geomodelling field, particularly in making geomodels more intelligent and adaptable to expert-driven interpretation, enabling multi-representation, and enhancing interoperability. The proposed approach offers potential for integration with other artificial intelligence disciplines, such as machine learning (ML). Several potential applications for incorporating ML into the geomodelling process have been identified, which present promising research directions. For instance, ML can be used to select inputs, providing an alternative approach to manual selection. The automation of input selection can be approached using feature selection algorithms or automated data pre-processing pipelines. Reinforcement learning (RL) algorithms can iteratively learn to improve input selection based on expert feedback or performance metrics. These models could be trained based on the outcomes of surveys on expert strategies or through self-learning techniques supervised by experts.

Furthermore, ML models could assist in evaluating interpretations or even generating new interpretations. For example, ML models trained on geologically constrained datasets could be used to perform specific interpretive tasks within the ‘PDCA’. Training these models could involve using large datasets that are consistent geologically and generated for instance through forward modeling, similar to the Noddyverse initiative by Jessel et al. (2022). Preliminary work in this area has been initiated through a graduate master thesis (report available upon request), and further developments are expected. Such approach can be combined with physics-informed neural networks (Cai et al., 2021) to consider the geomechincal aspects and deformation properties of geological objects.

Additionally, multimodal large language models (LLMs) could be explored. These LLMs, pre-trained on vast datasets from maps, reports, and scientific papers, could be fine-tuned with specific contextual data. The goal would be to integrate structured and formalized knowledge –such as knowledge graphs and ontological descriptions– during the training or fine-tuning phases. This would enable the inclusion of geological attributes such as lithologies, subsidence history, and deformation patterns, allowing for more accurate predictions of structural features by leveraging a high-dimensional latent space. The current research has already laid the groundwork for these applications, but further refinement and testing are needed.

At its end, I plead that the real deal in geology is not just having knowledge about rocks; it is about knowing how to rock this knowledge!

Bibliography

Abel, M., Gastal, E.S.L., Michelin, C.R.L., Maggi, L.G., Firnkes, B.E., Pachas, F.E.H., Alvarenga, R. dos S., 2019. A Knowledge Organization System for Image Classification and Retrieval in Petroleum Exploration Domain, in: Almeida, J.P.A., Bax, M., Berardi, R., Baião, F. (Eds.), Proceedings of the XII Seminar on Ontology Research in Brazil and III Doctoral and Masters Consortium on Ontologies, Porto Alegre, Brazil, September 2nd-5th, 2019, CEUR Workshop Proceedings. CEUR-WS.org.

Abel, M., Lorenzatti, A., Ros, L.F. de, Silva, O.P., Bernardes, A., Goldberg, K., Scherer, C.M., 2012a. Lithologic Logs in the Tablet through Ontology-based Facies Description, in: Abstract Volume AAPG Annual Convention & Exhibition. American Association of Petroleum Geologists, Tulsa, Oklahoma, USA.

Abel, M., Mastella, L.S., Perrin, M., Thonnat, M., 2012b. Ontologies and their use in geological knowledge formalization, in: Perrin, M., Rainaud, J.F. (Eds.), Shared Earth Modeling: Knowledge Based Solutions for Building and Managing Subsurface Structural Models. Technip, Paris, France, pp. 189–205.

Antoniou, G., Harmelen, F.V., 2009. Web Ontology Language: OWL, in: Staab, S., Studer, R. (Eds.), Handbook on Ontologies. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 91–110. https://doi.org/10.1007/978-3-540-92673-3_4

Arienti, G., Bistacchi, A., Caumon, G., Dal Piaz, G., Monopoli, B., Bertolo, D., 2024. Regional-scale 3D modelling in metamorphic belts: An implicit model-driven workflow applied in the Pennine Alps. Journal of Structural Geology 180, 105045. https://doi.org/10.1016/j.jsg.2023.105045

Armit, R.J., Betts, P.G., Schaefer, B.F., Ailleres, L., 2012. Constraints on long-lived Mesoproterozoic and Palaeozoic deformational events and crustal architecture in the northern Mount Painter Province, Australia. Gondwana Research 22, 207–226. https://doi.org/10.1016/j.gr.2011.11.003

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29. https://doi.org/10.1038/75556

Ashley, G.H., Cheney, M.G., Galloway, J.J., Gould, C.N., Hares, C.J., Howell, B.F., Levorsen, A.I., Miser, H.D., Moore, R.C., Reeside, J.B., Rubey, W.W., Stanton, T.W., Stose, G.W., Twenhofel, W.H., Representing the Association of American State Geologists, the United States Geological Survey, the American Association of Petroleum Geologists, and the Geological Society of America, on the Committee on Stratigraphic Nomenclature., 1933. Classification and Nomenclature of Rock Units. GSA Bulletin 44, 423–459. https://doi.org/10.1130/GSAB-44-423

Baader, F., Horrocks, I., Sattler, U., 2004. Description Logics, in: Staab, S., Studer, R. (Eds.), Handbook on Ontologies. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–28. https://doi.org/10.1007/978-3-540-24750-0_1

Baader, F., Nutt, W., 2003. Basic description logics, in: The Description Logic Handbook: Theory, Implementation, and Applications. pp. 43–95.

Babaie, H.A., Davarpanah, A., Banda, J.M., 2020. Ontology of Optical Microscopic Imaging and Analysis, in: AGU Fall Meeting Abstracts. pp. IN030-03.

Babaie, H.A., Oldow, J.S., Babaei, A., Lallemant, H.G.A., Watkinson, A.J., 2006. Designing a modular architecture for the structural geology ontology. Special Paper of the Geological Society of America 397, 269–282. https://doi.org/10.1130/2006.2397(21)

Bachimont, B., 2000. Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances. Ingénierie des connaissances: évolutions récentes et nouveaux défis 4, 305–323.

Baddeley, M.C., Curtis, A., Wood, R., 2004. An introduction to prior information derived from probabilistic judgements: elicitation of knowledge, cognitive bias and herding. Geological Society, London, Special Publications 239, 15–27. https://doi.org/10.1144/GSL.SP.2004.239.01.02

Bai, T., Tahmasebi, P., 2020. Hybrid geological modeling: Combining machine learning and multiple-point statistics. Computers and Geosciences 142, 104519. https://doi.org/10.1016/j.cageo.2020.104519

Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M.H., Bug, B., Chibucos, M.C., Clancy, K., Courtot, M., Derom, D., Dumontier, M., Fan, L., Fostel, J., Fragoso, G., Gibson, F., Gonzalez-Beltran, A., Haendel, M.A., He, Y., Heiskanen, M., Hernandez-Boussard, T., Jensen, M., Lin, Y., Lister, A.L., Lord, P., Malone, J., Manduchi, E., McGee, M., Morrison, N., Overton, J.A., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Schober, D., Smith, B., Soldatova, L.N., Stoeckert, C.J., Taylor, C.F., Torniai, C., Turner, J.A., Vita, R., Whetzel, P.L., Zheng, J., 2016. The Ontology for Biomedical Investigations. PLoS ONE 11, e0154556. https://doi.org/10.1371/journal.pone.0154556

Barwise, J., 1977. An Introduction to First-Order Logic, in: Studies in Logic and the Foundations of Mathematics. Elsevier, pp. 5–46. https://doi.org/10.1016/S0049-237X(08)71097-8

Bashore, W.M., Araktingi, U.G., Levy, M., Schweller, W.J., 1994. Importance of a Geological Framework and Seismic Data Integration for Reservoir Modeling and Subsequent Fluid-Flow Predictions, in: Yarus, J.M., Chambers, R.L. (Eds.), CA 3: Stochastic Modeling and Geostatistics, AAPG Special Volumes, Computer Applications 3. American Association of Petroleum Geologists, Tulsa, Oklahoma, USA.

Bastida, F., Aller, J., Fernández, F.J., Lisle, R.J., Bobillo-Ares, N.C., Menéndez, O., 2014. Recumbent folds: Key structural elements in orogenic belts. Earth-Science Reviews 135, 162–183. https://doi.org/10.1016/j.earscirev.2014.05.002

Belaid, N., Ait-Ameur, Y., Rainaud, J.-F., 2009. A semantic handling of geological modeling workflows, in: Proceedings of the International Conference on Management of Emergent Digital EcoSystems - MEDES '09. Presented at the the International Conference, ACM Press, France, p. 83. https://doi.org/10.1145/1643823.1643840

Belaid, N., Jean, S., Ait-Ameur, Y., Rainaud, J.-F., 2011. An ontology and indexation based management of services and workflows application to geological modeling. International Journal of Electronic Business Management 9, 296–309.

Belhachmi, A., 2024. An implicit spline-based method with PDE-based regularization for the construction of complex geological models (Theses). Université Côte d'Azur.

Belhachmi, A., Benabbou, A., Mourrain, B., 2024. A Spline-Based Regularized Method for the Reconstruction of Complex Geological Models. Math Geosci. https://doi.org/10.1007/s11004-024-10149-2

Bense, V.F., Gleeson, T., Loveless, S.E., Bour, O., Scibek, J., 2013. Fault zone hydrogeology. Earth-Science Reviews 127, 171–192. https://doi.org/10.1016/j.earscirev.2013.09.008

Berners-Lee, T., Hendler, J., Lassila, O., 2001. Web Semantic. Scientific American 284, 34–43.

Berners-Lee, T., Swick, R., 2006. Semantic Web Development (No. AFRL-IF-RS-TR-2006-294). Massachusetts Institute of Technology (MIT) / World Wide Web Consortium (W3C), Rome, New York.

Beven, K.J., Almeida, S., Aspinall, W.P., Bates, P.D., Blazkova, S., Borgomeo, E., Freer, J., Goda, K., Hall, J.W., Phillips, J.C., Simpson, M., Smith, P.J., Stephenson, D.B., Wagener, T., Watson, M., Wilkins, K.L., 2018. Epistemic uncertainties and natural hazard risk assessment – Part 1: A review of different natural hazard areas. Nat. Hazards Earth Syst. Sci. 18, 2741–2768. https://doi.org/10.5194/nhess-18-2741-2018

Boggs, S., 2006. Principles of sedimentology and stratigraphy, 4. ed. ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Bond, C.E., 2015. Uncertainty in structural interpretation: Lessons to be learnt. Journal of Structural Geology 74, 185–200. https://doi.org/10.1016/j.jsg.2015.03.003

Bond, C.E., Gibbs, A.D., Shipton, Z.K., Jones, S., 2007. What do you think this is? “Conceptual uncertainty” in geoscience interpretation. GSA Today 17, 4. https://doi.org/10.1130/GSAT01711A.1

Bond, C.E., Johnson, G., Ellis, J.F., 2015. Structural model creation: the impact of data type and creative space on geological reasoning and interpretation, in: Richards, F.L., Richardson, N.J., F. L. Richards, N. J. Richardson, S. J. Rippington, R. W. Wilson, BP and C. E. Bond (Eds.), Industrial Structural Geology: Principles, Techniques and Integration. Geological Society, London, Special Publications, pp. 83–97.

Bond, C.E., Lunn, R.J., Shipton, Z.K., Lunn, A.D., 2012. What makes an expert effective at interpreting seismic images? Geology 40, 75–78. https://doi.org/10.1130/G32375.1

Bond, C.E., Philo, C., Shipton, Z.K., 2011. When There isn’t a Right Answer: Interpretation and reasoning, key skills for twenty-first century geoscience. International Journal of Science Education 33, 629–652. https://doi.org/10.1080/09500691003660364

Bonnefoy, D., Chantraine, J., Lebret, P., Rabu, D., 1996. Carte géologique de la France à 1: 1000000 numérique. Rapport BRGM 39145.

Borgo, S., Ferrario, R., Gangemi, A., Guarino, N., Masolo, C., Porello, D., Sanfilippo, E.M., Vieu, L., 2022. DOLCE: A descriptive ontology for linguistic and cognitive engineering1. AO 17, 45–69. https://doi.org/10.3233/AO-210259

Bourgine, B., Prunier-Leparmentier, A.-M., Lembezat, C., Thierry, P., Luquet, C., Robelin, C., 2008. Tools and methods for constructing 3D geological models in the urban environment. The Paris case, in: Proceeding of the Eighth International Geostatistics Congress, JM Ortiz and X. Emery Editors, Vol2. pp. 951–960.

Brandel, S., Schneider, S., Perrin, M., Guiard, N., Rainaud, J.-F., Lienhard, P., Bertrand, Y., 2005. Automatic Building of Structured Geological Models. Journal of Computing and Information Science in Engineering 5, 138–148. https://doi.org/10.1115/1.1884145

Branquet, Y., Cheilletz, A., Giuliani, G., Laumonier, B., Blanco, O., 1999a. Fluidized hydrothermal breccia in dilatant faults during thrusting: the Colombian emerald deposits. Geo So SP 155, 183–195. https://doi.org/10.1144/GSL.SP.1999.155.01.14

Branquet, Y., Laumonier, B., Cheilletz, A., Giuliani, G., 1999b. Emeralds in the Eastern Cordillera of Colombia: Two tectonic settings for one mineralization. Geol 27, 597. https://doi.org/10.1130/0091-7613(1999)027<0597:EITECO>2.3.CO;2

Breitman, K.K., Casanova, M.A., Truszkowski, W., 2007. Semantic Web: concepts, technologies and applications, NASA monographs in systems and software engineering. Springer, New York.

Brodaric, B., 2012. Characterizing and representing inference histories in geologic mapping. International Journal of Geographical Information Science 26, 265–281. https://doi.org/10.1080/13658816.2011.585992

Brodaric, B., Gahegan, M., 2010. Ontology use for semantic e-Science. Semantic Web 1, 149–153. https://doi.org/10.3233/SW-2010-0021

Brodaric, B., Gahegan, M., Harrap, R., 2004. The art and science of mapping: computing geological categories from field data. Computers & Geosciences 30, 719–740. https://doi.org/10.1016/j.cageo.2004.05.001

Brodaric, B., Reitsma, F., Qiang, Y., 2008. SKIing with DOLCE: Toward an e-Science knowledge infrastructure, Formal Ontology in Information Systems, Proceedings of the Fifth International Conference (FOIS08). https://doi.org/10.3233/978-1-58603-923-3-208

Brodaric, B., Richard, S.M., 2020. The GeoScience Ontology 2020, IN030-07.

Burke, K., 2011. Plate Tectonics, the Wilson Cycle, and Mantle Plumes: Geodynamics from the Top. Annual Review of Earth and Planetary Sciences 39, 1–29. https://doi.org/10.1146/annurev-earth-040809-152521

Burke, K., Francis, P., Wells, G., 1990. Importance of the geological record in understanding global change. Global and Planetary Change 3, 193–204. https://doi.org/10.1016/0921-8181(90)90016-6

Cai, S., Mao, Z., Wang, Z., Yin, M., Karniadakis, G.E., 2021. Physics-informed neural networks (PINNs) for fluid mechanics: a review. Acta Mech. Sin. 37, 1727–1738. https://doi.org/10.1007/s10409-021-01148-1

Calcagno, P., Chilès, J.P., Courrioux, G., Guillen, A., 2008. Geological modelling from field data and geological knowledge. Physics of the Earth and Planetary Interiors 171, 147–157. https://doi.org/10.1016/j.pepi.2008.06.013

Calcagno, P., Courrioux, G., Guillen, A., Fitzgerald, D., McInerney, P., 2006. How 3D implicit geometric modelling helps to understand geology: the 3DGeoModeller methodology, in: Quantitative Geology from Multiple Sources. Presented at the XIth International Congress for Mathematical Geology, Liège, Belgium, p. 27.

Calcagno, P., Trumpy, E., Gutiérrez-Negrín, L.C., Liotta, D., 2022. A collection of 3D geomodels of the Los Humeros and Acoculco geothermal systems (Mexico). Sci Data 9, 280. https://doi.org/10.1038/s41597-022-01327-0

Carbonell, J.G., Michalski, R.S., Mitchell, T.M., 1983. An overview of Machine-Learning, in: Machine Learning. Elsevier, pp. 3–23. https://doi.org/10.1016/B978-0-08-051054-5.50005-4

Carbonera, J.L., Abel, M., Scherer, C.M., Bernardes, A.K., 2013. Visual Interpretation of Events in Petroleum Geology, in: 2013 IEEE 25th International Conference on Tools with Artificial Intelligence. Presented at the 2013 IEEE 25th International Conference on Tools with Artificial Intelligence (ICTAI), IEEE, Herndon, VA, USA, pp. 189–194. https://doi.org/10.1109/ICTAI.2013.37

Catuneanu, O., Galloway, W.E., Kendall, C.G.S. t. C., Miall, A.D., Posamentier, H.W., Strasser, A., Tucker, M.E., 2011. Sequence Stratigraphy: Methodology and Nomenclature. nos 44, 173–245. https://doi.org/10.1127/0078-0421/2011/0011

Caumon, G., 2010. Towards Stochastic Time-Varying Geological Modeling. Math Geosci 42, 555–569. https://doi.org/10.1007/s11004-010-9280-y

Caumon, G., 2003. Représentation, visualisation et modification de modèles volumiques pour les Géosciences (Theses). Institut National Polytechnique de Lorraine.

Caumon, G., Collon-Drouaillet, P., Le Carlier de Veslud, C., Viseur, S., Sausse, J., 2009. Surface-Based 3D Modeling of Geological Structures. Math Geosci 41, 927–945. https://doi.org/10.1007/s11004-009-9244-2

Caumon, G., Gray, G., Antoine, C., Titeux, M.-O., 2013. Three-Dimensional Implicit Stratigraphic Model Building From Remote Sensing Data on Tetrahedral Meshes: Theory and Application to a Regional Model of La Popa Basin, NE Mexico. IEEE Trans. Geosci. Remote Sensing 51, 1613–1621. https://doi.org/10.1109/TGRS.2012.2207727

Caumon, G., Lepage, F., Sword, C., Mallet, J.-L., Sword, C.H., 2004. Building and Editing a Sealed Geological Model Building and Editing a Sealed Geological Model 1. Mallet. Building and Editing a Sealed Geological Model. Mathematical Geology 36, 405–424. https://doi.org/10.1023/b:matg.0000029297.18098.8aï

Ceusters, W., Smith, B., 2015. Aboutness: Towards Foundations for the Information Artifact Ontology, in: Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO). CEUR vol. 1515, pp. 1–5.

Chambefort, I., Buscarlet, E., Wallis, I.C., Sewell, S., Wilmarth, M., 2016. Ngatamariki Geothermal Field, New Zealand: Geology, geophysics, chemistry and conceptual model. Geothermics 59, 266–280. https://doi.org/10.1016/j.geothermics.2015.07.011

Cherpeau, N., 2012. Incertitudes structurales en géomodélisation : échantillonnage et approche inverse (Thesis). Université de Lorraine.

Chilès, J.P., Aug, C., Guillén, A., Lees, T.C., 2007. Modelling the geometry of geological units and its uncertainty in 3D from structural data: The potential-field method, in: International Symposium on Orebody Modelling and Strategic Mine Planning. Presented at the International Symposium on Orebody Modelling and Strategic Mine Planning, Perth.

Cicconeto, F., Vieira, L.V., Abel, M., Alvarenga, R.D.S., Carbonera, J.L., Garcia, L.F., 2022. GeoReservoir: An ontology for deep-marine depositional system geometry description. Computers & Geosciences 159, 105005. https://doi.org/10.1016/j.cageo.2021.105005

Clary, R.M., Brzuszek, R.F., Wandersee, J.H., 2009. Students' Geocognition of Deep Time, Conceptualized in an Informal Educational Setting. Journal of Geoscience Education 57, 275–285. https://doi.org/10.5408/1.3544278

Coe, A.L. (Ed.), 2010. Geological field techniques. Wiley-Blackwell ; The Open University, Chichester, West Sussex, U.K.

Collon, P., Pichat, A., Kergaravat, C., Botella, A., Caumon, G., Ringenbach, J.-C., Callot, J.-P., 2016. 3D modeling from outcrop data in a salt tectonic context: Example from the Inceyol minibasin, Sivas Basin, Turkey. Interpretation 4, SM17–SM31. https://doi.org/10.1190/INT-2015-0178.1

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W.D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K., 2012. The SSN ontology of the W3C semantic sensor network incubator group. Journal of Web Semantics 17, 25–32. https://doi.org/10.1016/j.websem.2012.05.003

Courrioux, G., Allanic, C., Bourgine, B., Guillen, A., Baudin, T., Lacquement, F., Gabalda, S., Cagnard, F., Bayon, B.L., Besse, J., Marquer, D., Trap, P., Leloup, P.-H., Schreiber, D., 2015. Comparisons from multiple realizations of a geological model. Implication for uncertainty factors identification, in: IAMG 2015 : The 17th Annual Conference of the International Association for Mathematical Geosciences. IAMG (International Association for Mathematical Geosciences), Freiberg, Germany.

Cowan, E., Beatson, R., Fright, W., McLennan, T., Mitchell, T., 2002. Rapid geological modelling, in: Applied Structural Geology for Mineral Exploration and Mining, International Symposium. Kalgoorlie, pp. 23–25.

Cox, S.J.D., Richard, S.M., 2015. A geologic timescale ontology and service. Earth Sci Inform 8, 5–19. https://doi.org/10.1007/s12145-014-0170-6

Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Computers and Geosciences 63, 22–33. https://doi.org/10.1016/j.cageo.2013.10.008

d'Aquin, M., 2012. Modularizing Ontologies, in: Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E., Gangemi, A. (Eds.), Ontology Engineering in a Networked World. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 213–233. https://doi.org/10.1007/978-3-642-24794-1_10

Davies, J., 2010. Lightweight Ontologies, in: Poli, R., Healy, M., Kameas, A. (Eds.), Theory and Applications of Ontology: Computer Applications. Springer Netherlands, Dordrecht, pp. 197–229. https://doi.org/10.1007/978-90-481-8847-5_9

De Freitas, M.H., 2009. Geology; its principles, practice and potential for Geotechnics. QJEGH 42, 397–441. https://doi.org/10.1144/1470-9236/09-014

De Giacomo, G., Lenzerini, M., others, 1996. TBox and ABox reasoning in expressive description logics. KR 96, 10.

De Kemp, E., Jessell, M., Aillères, L., Schetselaar, E., Hillier, M., Lindsay, M., Brodaric, B., 2017. Earth model construction in challenging geologic terrain: Designing workflows and algorithms that makes sense, in: Proceedings of Exploration. pp. 419–439.

De Kemp, E.A., 2021. Spatial agents for geological surface modelling. Geosci. Model Dev. 14, 6661–6680. https://doi.org/10.5194/gmd-14-6661-2021

De Kemp, E.A., 1999. Visualization of complex geological structures using 3-D Bézier construction tools. Computers & Geosciences 25, 581–597. https://doi.org/10.1016/S0098-3004(98)00159-9

De La Varga, M., Wellmann, J.F., 2016. Structural geologic modeling as an inference problem: A Bayesian perspective. Interpretation 4, SM1–SM16. https://doi.org/10.1190/INT-2015-0188.1

De Laborda, C.P., Conrad, S., 2005. Querying Relational Databases with RDQL., in: Berliner XML Tage. Citeseer, pp. 161–172.
De Ville, B., 2013. Decision trees. WIREs Computational Stats 5, 448–455. https://doi.org/10.1002/wics.1278
Dell'Aversana, P., 2013. Cognition in geosciences: the feeding loop between geo-disciplines, cognitive sciences and epistemology. EAGE Publications, Netherlands.
Deming, W.E., 1997. The new economics for industry, government, education, 2. ed., 4. printing. ed. Massachusetts Institute of Technology / Center for Advanced Educational Services, Cambridge, MA.
Dermeval, D., Vilela, J., Bittencourt, I.I., Castro, J., Isotani, S., Brito, P., Silva, A., 2016. Applications of ontologies in requirements engineering: a systematic review of the literature. Requirements Eng 21, 405–437. https://doi.org/10.1007/s00766-015-0222-6
Di, H., Gao, D., 2016. Seismic attribute-aided fault detection in petroleum industry: A review, in: Fault Detection: Methods, Applications and Technology, Systems Engineering Methods, Developments and Technology. Nova Science Publishers, Inc, New York, NY, pp. 53–80.
Di, H., Li, C., Smith, S., Abubakar, A., 2020. Machine learning-assisted seismic interpretation with geologic constraints. SEG International Exposition and Annual Meeting 2019 5360–5364. https://doi.org/10.1190/segam2019-w4-01.1
Díaz, C., Dorner, B., Hussmann, H., Strijbos, J.-W., 2023. Conceptual review on scientific reasoning and scientific thinking. Curr Psychol 42, 4313–4325. https://doi.org/10.1007/s12144-021-01786-5
Dmitrieva, N.V., Nozhkin, A.D., 2012. Geochemistry of the Paleoproterozoic metaterrigenous rocks of the Biryusa Block, southwestern Siberian Craton. Lithol Miner Resour 47, 138–159. https://doi.org/10.1134/S0024490212010026
Doveton, J.H. (Ed.), 1994. Geologic Log Interpretation. SEPM (Society for Sedimentary Geology). https://doi.org/10.2110/scn.94.29
Dramsch, J.S., 2020. 70 Years of Machine Learning in Geoscience in Review. Advances in Geophysics 61, 1–55. https://doi.org/10.1016/bs.agph.2020.08.002
Drobnjakovic, M., Kulvatunyou, B., Ameri, F., Will, C., Smith, B., Jones, A., 2022. The Industrial Ontologies Foundry (IOF) Core Ontology, in: FOMI 2022: 12th International Workshop on Formal Ontologies Meet Industry, September 12-15, 2022, Tarbes, France. pp. 1–13.
Drummond, N., Shearer, R., 2006. The open world assumption, in: eSI Workshop: The Closed World of Databases Meets the Open World of the Semantic Web. p. 1.
Ekaputra, F.J., Serral, E., Winkler, D., Biffl, S., 2013. An analysis framework for ontology querying tools, in: Proceedings of the 9th International Conference on Semantic Systems. Presented at the ISEM '13: I-SEMANTICS 2013 - 9th International Conference on Semantic Systems, ACM, Graz Austria, pp. 1–8. https://doi.org/10.1145/2506182.2506183
Fabbri, P., Gaetan, C., Sartore, L., Dalla Libera, N., 2020. Subsoil Reconstruction in Geostatistics beyond Kriging: A Case Study in Veneto (NE Italy). Hydrology 7, 15. https://doi.org/10.3390/hydrology7010015
Falvey, D.A., 1974. The development of continental margins in plate tectonic theory. The APPEA Journal 14, 95. https://doi.org/10.1071/AJ73012
Faure, M., 2021. La chaine varisque en France, un édifice multi-collisionnel et poly-cyclique / Les quatre domaines varisques (1/2). Planet Terre.
Fauziati, S., Watanabe, K., 2010. Ontology of Volcano System and Volcanic Hazards Assessment. International Journal of Geoinformatics 6, 49.
Fernandez, O., Jones, S., Armstrong, N., Johnson, G., Ravaglia, A., Muñoz, J.A., 2009. Automated tools within workflows for 3D structural construction from surface and subsurface data. Geoinformatica 13, 291–304. https://doi.org/10.1007/s10707-008-0059-y
Fernández-López, M., Gómez-Pérez, A., 2002. Overview and analysis of methodologies for building ontologies. The Knowledge Engineering Review 17, 129–156. https://doi.org/10.1017/S0269888902000462

Fikes, R., Hayes, P., Horrocks, I., 2004. OWL-QL—a language for deductive query answering on the Semantic Web. Journal of Web Semantics 2, 19–29. https://doi.org/10.1016/j.websem.2004.07.002

Forsythe, D.E., 1993. Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. Soc Stud Sci 23, 445–477. https://doi.org/10.1177/030631293023003002

Fossen, H., 2016. Structural geology. Cambridge university press.

Foxvog, D., 2010. Cyc, in: Poli, R., Healy, M., Kameas, A. (Eds.), Theory and Applications of Ontology: Computer Applications. Springer Netherlands, Dordrecht, pp. 259–278. https://doi.org/10.1007/978-90-481-8847-5_12

Franke, R., 1982. Scattered data interpolation: tests of some methods. Mathematics of Computation 38, 181–200. https://doi.org/10.1090/s0025-5718-1982-0637296-4

Fredrich, J.T., Coblentz, D., Fossum, A.F., Thorne, B.J., 2003. Stress Perturbations Adjacent to Salt Bodies in the Deepwater Gulf of Mexico, in: All Days. Presented at the SPE Annual Technical Conference and Exhibition, SPE, Denver, Colorado, p. SPE-84554-MS. https://doi.org/10.2118/84554-MS

Frodeman, R., 2014. Hermeneutics in the Field: The Philosophy of Geology, in: Babich, B., Ginev, D. (Eds.), The Multidimensionality of Hermeneutic Phenomenology, Contributions to Phenomenology. Springer International Publishing, Cham, pp. 69–79. https://doi.org/10.1007/978-3-319-01707-5_5

Frodeman, R., 1995. Geological reasoning: Geology as an interpretive and historical science. Geological Society of America Bulletin, 107, 960–968.

Fry, N., 1997. The field description of metamorphic rocks, Repr. ed, Geological Society of London Handbook. Wiley, Chichester.

Galley, C.G., Lelièvre, P.G., Farquharson, C.G., 2020. Geophysical inversion for 3D contact surface geometry. Geophysics 85, K27–K45. https://doi.org/10.1190/geo2019-0614.1

Garcia, L.F., Abel, M., Perrin, M., Alvarenga, R. dos S., 2020. The GeoCore ontology: A core ontology for general use in Geology. Computers and Geosciences 135, 104387. https://doi.org/10.1016/j.cageo.2019.104387

Geman, S., Geman, D., 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-6, 721–741. https://doi.org/10.1109/TPAMI.1984.4767596

Gill, R., Fitton, J.G., 2022. Igneous rocks and processes: a practical guide, Second edition. ed. Wiley, Hoboken, NJ.

Giraud, J., Jessell, M., Lindsay, M., Parkyuz-Charrier, E., Martin, R., 2016. Integrated geophysical joint inversion using petrophysical constraints and geological modelling. Presented at the 2016 SEG International Exposition and Annual Meeting, OnePetro.

Giunchiglia, F., Zaihrayeu, I., 2007. Lightweight Ontologies (Departmental Technical Report No. DIT-07-071). Università degli Studi di Trento, Information Engineering and Computer Science Department.

Glimm, B., Horrocks, I., Motik, B., Stoilos, G., Wang, Z., 2014. HermiT: An OWL 2 Reasoner. J Autom Reasoning 53, 245–269. https://doi.org/10.1007/s10817-014-9305-1

Godefroy, G., Caumon, G., Laurent, G., Bonneau, F., 2019. Structural Interpretation of Sparse Fault Data Using Graph Theory and Geological Rules: Fault Data Interpretation. Mathemathical Geosciences 51, 1091–1107. https://doi.org/10.1007/s11004-019-09800-0

Gómez-Pérez, A., 1999. Ontological engineering: A state of the art. Expert Update: Knowledge Based Systems and Applied Artificial Intelligence 2, 33–43.

Grose, L., Ailleres, L., 2021. Inversion of geological maps for fault geometry, in: 2021 RING Meeting. ASGA.

Grose, L., Ailleres, L., Laurent, G., Armit, R., Jessell, M., 2019. Inversion of geological knowledge for fold geometry. Journal of Structural Geology 119, 1–14. https://doi.org/10.1016/j.jsg.2018.11.010

Grose, L., Ailleres, L., Laurent, G., Caumon, G., Jessell, M., Armit, R., 2021a. Realistic modelling of faults in LoopStructural 1.0. https://doi.org/10.5194/gmd-2021-112

Grose, L., Aillères, L., Laurent, G., Jessell, M., 2021b. LoopStructural 1.0: time-aware geological modelling. Geoscientific Model Development 14, 3915–3937. https://doi.org/10.5194/gmd-14-3915-2021

Grose, L., Laurent, G., Aillères, L., Armit, R., Jessell, M., Caumon, G., 2017. Structural data constraints for implicit modeling of folds. Journal of Structural Geology 104, 80–92. https://doi.org/10.1016/j.jsg.2017.09.013

Grose, L., Laurent, G., Aillères, L., Armit, R., Jessell, M., Cousin-Dechenaud, T., 2018. Inversion of Structural Geology Data for Fold Geometry. Journal of Geophysical Research: Solid Earth 123, 6318–6333. https://doi.org/10.1029/2017JB015177

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies 43, 907–928. https://doi.org/10.1006/ijhc.1995.1081

Guarino, N., 1997. Understanding, building and using ontologies. International Journal of Human-Computer Studies 46, 293–310. https://doi.org/10.1006/ijhc.1996.0091

Guarino, N., Giaretta, P., 1995. Ontologies and knowledge bases, in: Mars, N.J.I. (Ed.), Towards Very Large Knowledge Bases : Knowledge Building & Knowledge Sharing. IOS Press, Amsterdam, pp. 25–32.

Guarino, N., Guizzardi, G., Mylopoulos, J., 2019. On the Philosophical Foundations of Conceptual Models.

Guarino, N., Welty, C.A., 2009. An Overview of OntoClean, in: Staab, S., Studer, R. (Eds.), Handbook on Ontologies. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 201–220. https://doi.org/10.1007/978-3-540-92673-3_9

Guillen-Rondon, P., Cobos, C., Larrazabal, G., Diz, A., 2020. Machine learning: A deep learning approach for seismic structural evaluation. SEG International Exposition and Annual Meeting 2019 4990–4994. https://doi.org/10.1190/segam2019-3216712.1

Guizzardi, G., 2007. On ontology, ontologies, conceptualizations, modeling languages, and (meta)models. Frontiers in Artificial Intelligence and Applications 155, 18–39.

Guizzardi, G., 2005. Ontological foundations for structural conceptual models (PhD Thesis - Research UT, graduation UT). CTIT PhD Thesis Series. University of Twente.

Guizzardi, G., Wagner, G., 2010. Using the Unified Foundational Ontology (UFO) as a Foundation for General Conceptual Modeling Languages, in: Poli, R., Healy, M., Kameas, A. (Eds.), Theory and Applications of Ontology: Computer Applications. Springer Netherlands, Dordrecht, pp. 175–196. https://doi.org/10.1007/978-90-481-8847-5_8

Gumiaux, C., 2003. Modélisation du cisaillement hercynien de Bretagne centrale: déformation crustale et implications lithosphériques (Thesis). Université Rennes 1.

Gunderson, K.L., Zhang, Z., Payne, B., Cheng, S., Jiang, Z., Wang, A., 2022. Machine learning applications to seismic structural interpretation: Philosophy, progress, pitfalls, and potential. AAPG Bulletin 106, 2187–2202. https://doi.org/10.1306/12162121016

Guo, J., Wang, J., Wu, L., Liu, C., Li, C., Li, F., Lin, M., Jessell, M.W., Li, P., Dai, X., Tang, J., 2020. Explicit-implicit-integrated 3-D geological modelling approach: A case study of the Xianyan Demolition Volcano (Fujian, China). Tectonophysics 795, 228648. https://doi.org/10.1016/j.tecto.2020.228648

Haller, A., Polleres, A., 2020. Are we better off with just one ontology on the Web? Semantic Web 11, 87–99.

Hardy, R.L., 1971. Multiquadric equations of topography and other irregular surfaces. J. Geophys. Res. 76, 1905–1915. https://doi.org/10.1029/JB076i008p01905

Hasan, M., Shang, Y., 2022. Geophysical evaluation of geological model uncertainty for infrastructure design and groundwater assessments. Engineering Geology 299, 106560. https://doi.org/10.1016/j.enggeo.2022.106560

He, K., Zhang, Z., Dong, Y., Cai, D., Lu, Y., Han, W., 2024. Improving Geological Remote Sensing Interpretation via a Contextually Enhanced Multiscale Feature Fusion Network. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 1–17. https://doi.org/10.1109/JSTARS.2024.3374818

He, X., Højberg, A.L., Jørgensen, F., Refsgaard, J.C., 2015. Assessing hydrological model predictive uncertainty using stochastically generated geological models: Hydrological Modelling Using Stochastic Geological Models. Hydrol. Process. 29, 4293–4311. https://doi.org/10.1002/hyp.10488

Heather R. Miller, K.S.M., Herbert, B.E., 2010. Inquiry in the Physical Geology Classroom: Supporting Students' Conceptual Model Development. Journal of Geography in Higher Education 34, 595–615. https://doi.org/10.1080/03098265.2010.499562

Heim, E., Virgo, S., de la Varga, M., Wellmann, F., 2019. Is it possible to capture epistemic uncertainty in geological models?, in: Geophysical Research Abstracts. Presented at the EGU General Assembly 2019, Vienna.

Helton, J.C., Sallaberry, C.J., 2017. 17 - Treatment of uncertainty in performance assessments for the geological disposal of radioactive waste, in: Apted, M.J., Ahn, J. (Eds.), Geological Repository Systems for Safe Disposal of Spent Nuclear Fuels and Radioactive Waste (Second Edition), Woodhead Publishing Series in Energy. Woodhead Publishing, pp. 499–527. https://doi.org/10.1016/B978-0-08-100642-9.00017-7

Hillier, M., Wellmann, F., De Kemp, E.A., Brodaric, B., Schetselaar, E., Bédard, K., 2023. GeoINR 1.0: an implicit neural network approach to three-dimensional geological modelling. Geosci. Model Dev. 16, 6987–7012. https://doi.org/10.5194/gmd-16-6987-2023

Hillier, M.J., Schetselaar, E.M., de Kemp, E.A., Perron, G., 2014. Three-Dimensional Modelling of Geological Surfaces Using Generalized Interpolation with Radial Basis Functions. Mathematical Geosciences 46, 931–953. https://doi.org/10.1007/s11004-014-9540-3

Hills, E.S., 1970. Elements of Structural Geology. Springer Science & Business Media.

Himus, B.G.W., Sweeting, G.S., 1952. The Elements of Field Geology. Geological Magazine 89, 382–382. https://doi.org/10.1017/S0016756800068047

Hollis D. Hedberg (2), 1958. Stratigraphic Classification and Terminology. Bulletin 42. https://doi.org/10.1306/0BDA5B85-16BD-11D7-8645000102C1865D

Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M., others, 2004. SWRL: A semantic web rule language combining OWL and RuleML. W3C Member submission 21, 1–31.

Howard, A.S., Hatton, B., Reitsma, F., Lawrie, K.I.G., 2009. Developing a geoscience knowledge framework for a national geological survey organisation. Computers & Geosciences 35, 820–835. https://doi.org/10.1016/j.cageo.2008.06.004

Huang, J.-C., Li, A.-B., Wang, X., Shao, C.-Z., Shen, Y.-G., 2023. Automated identification and mapping of geological folds in cross sections. Open Geosciences 15, 20220479. https://doi.org/10.1515/geo-2022-0479

Iglesias, M., Lu, Y., Stuart, A., 2016. A Bayesian level set method for geometric inverse problems. Interfaces Free Bound. 18, 181–217. https://doi.org/10.4171/IFB/362

Ivanović, M., Budimac, Z., 2014. An overview of ontologies and data resources in medical domains. Expert Systems with Applications 41, 5158–5166. https://doi.org/10.1016/j.eswa.2014.02.045

Jackson, M.P.A., Vendeville, B.C., Schultz-Ela, D.D., 1994. Structural dynamics of salt systems. Annu. Rev. Earth Planet. Sci. 22, 93–117. https://doi.org/10.1146/annurev.ea.22.050194.000521

Jaffri, A., Glaser, H., Millard, I., 2007. URI Identity Management for Semantic Web Data Integration and Linkage, in: Meersman, R., Tari, Z., Herrero, P. (Eds.), On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1125–1134. https://doi.org/10.1007/978-3-540-76890-6_40

Jayr, S., Gringarten, E., Tertois, A.L., Mallet, J.L., Dulac, J.C., 2008. The need for a correct geological modelling support: the advent of the UVT-transform. First Break 26. https://doi.org/10.3997/1365-2397.26.10.28558

Jean, S., 2007. OntoQL, un langage d'exploitation des bases de données à base ontologique (Theses). Université de Poitiers.

Jessell, M., Guo, J., Li, Y., Lindsay, M., Scalzo, R., Giraud, J., Pirot, G., Cripps, E., Ogarko, V., 2022. Into the Noddyverse: a massive data store of 3D geological models for machine learning and inversion applications. Earth System Science Data 14, 381–392. https://doi.org/10.5194/essd-14-381-2022

Jessell, M., Pakyuz-Charrier, E., Lindsay, M., Giraud, J., De Kemp, E., 2018. Assessing and Mitigating Uncertainty in Three-Dimensional Geologic Models in Contrasting Geologic Scenarios, in: Arribas R., A.M., Mauk, J.L. (Eds.), Metals, Minerals, and Society. Society of Economic Geologists, pp. 63–74. https://doi.org/10.5382/SP.21.04

Jessell, M.W., Valenta, R.K., 1996. Structural geophysics: Integrated structural and geophysical modelling, in: G. De Paor, D. (Ed.), Computer Methods in the Geosciences, Structural Geology and Personal Computers. Pergamon Press, Oxford, pp. 303–324.

Jiang, Y., Tang, Y., Chen, Q., Wang, J., Tang, S., 2010. Extending soft sets with description logics. Computers & Mathematics with Applications 59, 2087–2096. https://doi.org/10.1016/j.camwa.2009.12.014

Jørgensen, F., Høyer, A.-S., Sandersen, P.B.E., He, X., Foged, N., 2015. Combining 3D geological modelling techniques to address variations in geology, data type and density – An example from Southern Denmark. Computers & Geosciences 81, 53–63. https://doi.org/10.1016/j.cageo.2015.04.010

Kalemi, E., Martiri, E., 2011. FOAF-Academic Ontology: A Vocabulary for the Academic Community, in: 2011 Third International Conference on Intelligent Networking and Collaborative Systems. Presented at the 2011 Third International Conference on Intelligent Networking and Collaborative Systems (INCoS), IEEE, Fukuoka, Japan, pp. 440–445. https://doi.org/10.1109/INCoS.2011.94

Kang, B., Lee, K., 2020. Managing uncertainty in geological scenarios using machine learning-based classification model on production data. Geofluids 2020. https://doi.org/10.1155/2020/8892556

Karray, M., Otte, N., Rai, R., Ameri, F., Kulvatunyou, B., Smith, B., Kiritsis, D., Will, C., Arista, R., Others, 2021. The Industrial Ontologies Foundry (Iof) Perspectives, in: Proceedings: Industrial Ontology Foundry (IOF) Achieving Data Interoperability Workshop, International Conference on Interoperability for Enterprise Systems and Applications, Tarbes, France, March 17-24, 2020.

Kastens, K.A., Ishikawa, T., 2006. Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields, in: Earth and Mind: How Geologists Think and Learn about the Earth. Geological Society of America. https://doi.org/10.1130/2006.2413(05)

Katsumi, M., Grüninger, M., 2016. What is ontology reuse?, in: Formal Ontology in Information Systems. IOS Press, pp. 9–22.

Kessler, H., Mathers, S., Sobisch, H.G., 2009. The capture and dissemination of integrated 3D geospatial knowledge at the British Geological Survey using GSI3D software and methodology. Computers and Geosciences 35, 1311–1321. https://doi.org/10.1016/j.cageo.2008.04.005

Khaninezhad, M.M., Jafarpour, B., Li, L., 2012. Sparse geologic dictionaries for subsurface flow model calibration: Part II. Robustness to uncertainty. Advances in Water Resources 39, 122–136. https://doi.org/10.1016/j.advwatres.2011.10.005

Kolesov, V., 2018. Cognitive Modelling in Oil & Gas Exploration and Reservoir Prediction. Presented at the 80th EAGE Conference and Exhibition 2018, European Association of Geoscientists & Engineers, pp. 1–5. https://doi.org/10.3997/2214-4609.201800905

Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E., 2006. Machine learning: a review of classification and combining techniques. Artif Intell Rev 26, 159–190. https://doi.org/10.1007/s10462-007-9052-3

Krajnovich, A., Zhou, W., Gutierrez, M., 2020. Uncertainty assessment for 3D geologic modeling of fault zones based on geologic inputs and prior knowledge. Solid Earth 11, 1457–1474. https://doi.org/10.5194/se-11-1457-2020

Kuhn, D., 2010. What is scientific thinking and how does it develop? The Wiley-Blackwell handbook of childhood cognitive development 497–523.

Kuhn, T.S., 1997. The structure of scientific revolutions. University of Chicago press Chicago.

Lacchia, A., Schuitema, G., McAuliffe, F., 2020. The human side of geoscientists: comparing geoscientists' and non-geoscientists' cognitive and affective responses to geology. Geoscience Communication 3, 291–302. https://doi.org/10.5194/gc-3-291-2020

Lajaunie, C., Courrioux, G., Manuel, L., 1997. Foliation fields and 3D cartography in geology: Principles of a method based on potential interpolation. Mathematical Geology 29, 571–584. https://doi.org/10.1007/bf02775087

Lamotte, D.F. de, Leturmy, P., Souloumiac, P., Lamotte, A.F. de, 2020. Geological Objects and Structures in 3D: Observation, Interpretation and Building of 3D Models. CRC Press, London. https://doi.org/10.1201/9781003047230

Lamy, J.-B., 2021. Ontologies with Python Programming OWL 2.0 Ontologies with Python and Owlready2, 1st ed, Ontologies with Python. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6552-9

Laouici, I., Laurent, G., Loiselet, C., Branquet, Y., 2024. A knowledge-driven modeling formalism for automatic structural interpretation.

Laurent, G., 2016. Iterative thickness regularization of stratigraphic layers in discrete implicit modeling. Mathematical Geosciences 48, 811–833.

Laurent, G., 2013. Prise en compte de l'histoire géologique des structures dans la création de modèles numériques 3D compatibles (Thesis). University of Lorraine.

Laurent, G., Ailleres, L., Grose, L., Caumon, G., Jessell, M., Armit, R., 2016. Implicit modeling of folds and overprinting deformation. Earth and Planetary Science Letters 456, 26–38. https://doi.org/10.1016/j.epsl.2016.09.040

Laurent, G., Caumon, G., Bouziat, A., Jessell, M., 2013. A parametric method to model 3D displacements around faults with volumetric vector fields. Tectonophysics 590, 83–93. https://doi.org/10.1016/j.tecto.2013.01.015

Le Maitre, R.W. (Ed.), 2004. Igneous rocks: a classification and glossary of terms: recommendations of the International Union of Geological Sciences, Subcommission on the Systematics of Igneous Rocks, 2nd ed. ed, International Union of Geological Sciences. Cambridge University Press, Cambridge ; New York.

Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., Zhao, J., 2013. PROV-O: The PROV Ontology.

Lecue, F., 2020. On the role of knowledge graphs in explainable AI. SW 11, 41–51. https://doi.org/10.3233/SW-190374

Legg, C.A., King, M.J., Deny, L., Rainaud, J.-F., Hollingsworth, J., Marcotte, D., 2015. Representing Subsurface Knowledge Using RESQML. Presented at the SPE Digital Energy Conference and Exhibition, OnePetro.

Lehmann, F., 1992. Semantic networks. Computers & Mathematics with Applications 23, 1–50.

Lemon, A.M., Jones, N.L., 2003. Building solid models from boreholes and user-defined cross-sections. Computers & Geosciences 29, 547–555. https://doi.org/10.1016/S0098-3004(03)00051-7

Levens, R.L., Williams, R.E., Ralston, D.R., 1994. Hydrogeologic role of geologic structures. Part 1: the paradigm. Journal of Hydrology 156, 227–243.

Levy, B., Mallet, J.-L., 1998. Non-Distorted Texture Mapping For Sheared Triangulated Meshes, in: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques. Association for Computing Machinery, New York, United States, pp. 343–352.

Liang, D., Hua, W., Liu, X., Zhao, Y., Liu, Z., 2021. Uncertainty assessment of a 3D geological model by integrating data errors, spatial variations and cognition bias. Earth Science Informatics 14, 161–178.

Libarkin, J., 2014. The Role of Scholarly Publication in Geocognition and Discipline-Based Geoscience Education Research, in: Tong, V.C.H. (Ed.), Geoscience Research and Education: Teaching at Universities. Springer Netherlands, Dordrecht, pp. 69–76. https://doi.org/10.1007/978-94-007-6946-5_7

Libarkin, J., 2009. Geocognition research: an international discipline, in: AGU Fall Meeting Abstracts. pp. ED44A-03.

Liben, L.S., Christensen, A.E., Kastens, K.A., 2010. Gestures in Geology: The Roles of Spatial Skills, Expertise, and Communicative Context, in: Hölscher, C., Shipley, T.F., Olivetti Belardinelli, M., Bateman, J.A., Newcombe, N.S. (Eds.), Spatial Cognition VII, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 95–111. https://doi.org/10.1007/978-3-642-14749-4_11

Lidal, E.M., Hauser, H., Viola, I., 2012. Geological Storytelling - Graphically Exploring and Communicating Geological Sketches. Eurographics Workshop on Sketch-Based Interfaces and Modeling 10. https://doi.org/10.2312/SBM/SBM12/011-020

Lidal, E.M., Natali, M., Patel, D., Hauser, H., Viola, I., 2013. Geological storytelling. Computers & Graphics 37, 445–459. https://doi.org/10.1016/j.cag.2013.01.010

Loiselet, C., Bellier, C., Lopez, S., COURRIOUX, C., 2016. Storing and delivering numerical geological models on demand for everyday Earth Sciences applications, in: 35th International Geological Congress. Cape Town, South Africa.

Lombardo, V., Piana, F., Mimmo, D., 2018. Semantics–informed geological maps: Conceptual modeling and knowledge encoding. Computers and Geosciences 116, 12–22. https://doi.org/10.1016/j.cageo.2018.04.001

López, G.I., 2015. Walther's Law of Facies, in: Jack Rink, W., Thompson, J.W. (Eds.), Encyclopedia of Scientific Dating Methods, Encyclopedia of Earth Sciences Series. Springer Netherlands, Dordrecht, pp. 957–958. https://doi.org/10.1007/978-94-007-6304-3_30

Lyu, M., Ren, B., Wu, B., Tong, D., Ge, S., Han, S., 2021. A parametric 3D geological modeling method considering stratigraphic interface topology optimization and coding expert knowledge. Engineering Geology 293, 106300. https://doi.org/10.1016/j.enggeo.2021.106300

Ma, X., Carranza, E.J.M., Wu, C., Meer, F.D.V. der, 2012. Ontology-aided annotation, visualization, and generalization of geological time-scale information from online geological map services. Computers and Geosciences 40, 107–119. https://doi.org/10.1016/j.cageo.2011.07.018

Madani Esfahani, N., Asghari, O., 2013. Fault detection in 3D by sequential Gaussian simulation of Rock Quality Designation (RQD): Case study: Gazestan phosphate ore deposit, Central Iran. Arab J Geosci 6, 3737–3747. https://doi.org/10.1007/s12517-012-0633-3

Maeng, S., Lee, K., 2015. Cross-sectional item response analysis of geocognition assessment for the development of plate tectonics learning progressions: Rasch model. Journal of the Korean Association for Science Education 35, 37–52.

Maljers, D., Stafleu, J., Van Der Meulen, M.J., Dambrink, R.M., 2015. Advances in constructing regional geological voxel models, illustrated by their application in aggregate resource assessments. Netherlands Journal of Geosciences 94, 257–270. https://doi.org/10.1017/njg.2014.46

Mallet, J.-L., 2002. Geomodeling. Oxford University Press.

Mallet, J.L., 1997. Discrete modeling for natural objects. Math Geol 29, 199–219. https://doi.org/10.1007/BF02769628

Mallet, J.L., 1992. GOCAD: A Computer Aided Design Program for Geological Applications, in: Turner, A.K. (Ed.), Three-Dimensional Modeling with Geoscientific Information Systems. Springer Netherlands, Dordrecht, pp. 123–141. https://doi.org/10.1007/978-94-011-2556-7_11
Mallet, J.L., Jacquemin, P., Cheimanoff, N., 1989. GOCAD project: Geometric modeling of complex geological surfaces. SEG Technical Program Expanded Abstracts 1989. https://doi.org/10.1190/1.1889515
Manduca, C.A., Mogk, D.W., 2006. Earth and Mind: How Geologists Think and Learn about the Earth, Special Paper 413: Earth and Mind: How Geologists Think and Learn about the Earth. https://doi.org/10.1130/0-8137-2413-9
Manktelow, K.I., 1999. Reasoning and thinking. Psychology press.
Mantovani, A., Piana, F., Lombardo, V., 2020. Ontology-driven representation of knowledge for geological maps. Computers and Geosciences 139, 104446. https://doi.org/10.1016/j.cageo.2020.104446
Marinoni, O., 2003. Improving geological models using a combined ordinary–indicator kriging approach. Engineering Geology 69, 37–45. https://doi.org/10.1016/S0013-7952(02)00246-6
Mas, P., Calcagno, P., Caritg-Monnot, S., Beccaletto, L., Capar, L., Hamm, V., 2022. A 3D geomodel of the deep aquifers in the Orléans area of the southern Paris Basin (France). Sci Data 9, 781. https://doi.org/10.1038/s41597-022-01876-4
Mastella, L.S., Ait-Ameury, Y., Jean, S., Perrin, M., Rainaud, J.-F., 2009. Semantic exploitation of persistent metadata in engineering models: application to geological models, in: 2009 Third International Conference on Research Challenges in Information Science. IEEE, pp. 129–138.
Maxelon, M., Renard, P., Courrioux, G., Brändli, M., Mancktelow, N., 2009. A workflow to facilitate three-dimensional geometrical modelling of complex poly-deformed geological units. Computers & Geosciences 35, 644–658. https://doi.org/10.1016/j.cageo.2008.06.005
Mazzetti, P., Nativi, S., Santoro, M., Giuliani, G., Rodila, D., Folino, A., Caruso, S., Aracri, G., Lehmann, A., 2022. Knowledge formalization for Earth Science informed decision-making: The GEOEssential Knowledge Base. Environmental Science & Policy 131, 93–104. https://doi.org/10.1016/j.envsci.2021.12.023
McInerney, P., Guillen, A., Courrioux, G., Calcagno, P., Lees, T., 2005. Building 3D Geological Models Directly from the Data? A New Approach Applied to Broken Hill, Australia.
Mei, G., 2014. Summary on Several Key Techniques in 3D Geological Modeling. The Scientific World Journal 2014, 1–11. https://doi.org/10.1155/2014/723832
Michon, L., Merle, O., 2003. Mode of lithospheric extension: Conceptual models from analogue modeling. Tectonics 22, 2002TC001435. https://doi.org/10.1029/2002TC001435
Middendorp, C.P., 1991. On the conceptualization of theoretical constructs. Qual Quant 25, 235–252. https://doi.org/10.1007/BF00167530
Mishra, R.B., Kumar, S., 2011. Semantic web reasoners and languages. Artif Intell Rev 35, 339–368. https://doi.org/10.1007/s10462-010-9197-3
Moen, R., 2009. Foundation and History of the PDSA Cycle, in: Asian Network for Quality Conference. Tokyo. Https://Www. Deming. Org/Sites/Default/Files/Pdf/2015/PDSA_History_Ron_Moen. Pdf.
Mogk, D.W., Goodwin, C., 2012. Learning in the field: Synthesis of research on thinking and learning in the geosciences, in: Earth and Mind II: A Synthesis of Research on Thinking and Learning in the Geosciences. Geological Society of America. https://doi.org/10.1130/2012.2486(24)
Mohri, M., Rostamizadeh, A., Talwalkar, A., 2012. Foundations of machine learning, Adaptive computation and machine learning series. MIT Press, Cambridge, MA.
Monfardini, G.K.Q., Salamon, J.S., Barcellos, M.P., 2023. Use of Competency Questions in Ontology Engineering: A Survey, in: Almeida, J.P.A., Borbinha, J., Guizzardi, G., Link, S., Zdravkovic, J. (Eds.), Conceptual Modeling, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 45–64. https://doi.org/10.1007/978-3-031-47262-6_3
Morandini, F., Michel, B., Verney, P., Rainaud, J.-F., Deny, L., Dulac, J.-C., Fitzpatrick, T., Eastick, R., Towery, L., 2011. Using RESQML for Shared Earth Model Data Exchanges between

Commercial Modelling Applications and In-House Developments, Demonstrated on Actual Subsurface Data, in: All Days. Presented at the SPE Digital Energy Conference and Exhibition, SPE, The Woodlands, Texas, USA, p. SPE-143846-MS. https://doi.org/10.2118/143846-MS
Möystad, O., 2017. Cognition and the Built Environment, Cognition and the Built Environment. Taylor and Francis. https://doi.org/10.4324/9781315642383
Myers, J.S., 1975. Igneous stratigraphy of Archaean anorthosite at Majorqap qâva, near Fiskenæsset, South-West Greenland. rapggu 74, 1–27. https://doi.org/10.34194/rapggu.v74.7431
Myerson, R.B., 2013. Game Theory: Analysis of Conflict. Harvard University Press. https://doi.org/10.2307/j.ctvjsf522
Nabavi, S.T., Fossen, H., 2021. Fold geometry and folding – a review. Earth-Science Reviews 222, 103812. https://doi.org/10.1016/j.earscirev.2021.103812
Nardi, D., Brachman, R.J., 2007. An Introduction to Description Logics, in: Baader, F., Calvanese, D., McGuinness, D.L., Nardi, Daniele, Patel-Schneider, P.F. (Eds.), The Description Logic Handbook. Cambridge University Press, pp. 1–44. https://doi.org/10.1017/CBO9780511711787.003
Newell, A., Simon, H.A., others, 1972. Human problem solving. Prentice-hall Englewood Cliffs, NJ.
Newell, A., Tonge, F.M., 1960. An introduction to information processing language V. Communications of the ACM 3, 205–211.
Niles, I., Pease, A., 2003. Mapping WordNet to the SUMO ontology, in: Proceedings of the Ieee International Knowledge Engineering Conference. Citeseer, pp. 23–26.
Noy, N.F., Crubézy, M., Fergerson, R.W., Knublauch, H., Tu, S.W., Vendetti, J., Musen, M.A., 2003. Protégé-2000: an open-source ontology-development and knowledge-acquisition environment., in: AMIA... Annual Symposium Proceedings. AMIA Symposium. pp. 953–953.
Oakley, D., Loiselet, C., Coowar, T., Labbe, V., Callot, J.-P., 2024. GEOMAPLEARN 1.0: Detecting geological structures from geological maps with machine learning. https://doi.org/10.5194/gmd-2024-35
O'Connell, S.C., Dix, G.R., Barclay, J.E., 1990. The origin, history, and regional structural development of the Peace River Arch, Western Canada. Bulletin of Canadian Petroleum Geology 38A, 4–24. https://doi.org/10.35767/gscpgbull.38a.1.004
Olierook, H.K.H., Scalzo, R., Kohn, D., Chandra, R., Farahbakhsh, E., Clark, C., Reddy, S.M., Müller, R.D., 2021. Bayesian geological and geophysical data fusion for the construction and uncertainty quantification of 3D geological models. Geoscience Frontiers 12, 479–493. https://doi.org/10.1016/j.gsf.2020.04.015
Oliver, M.A., Webster, R., 1990. Kriging: a method of interpolation for geographical information systems. International journal of geographical information systems 4, 313–332. https://doi.org/10.1080/02693799008941549
Olszewski Jr., W.J., 1980. The geochronology of some stratified metamorphic rocks in northeastern Massachusetts. Can. J. Earth Sci. 17, 1407–1416. https://doi.org/10.1139/e80-148
Otte, J.N., Beverley, J., Ruttenberg, A., 2022. BFO: Basic Formal Ontology1. AO 17, 17–43. https://doi.org/10.3233/AO-220262
Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J., Jessell, M., 2018. Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modeling, a guide for disturbance distribution selection and parameterization. Solid Earth 9, 385–402. https://doi.org/10.5194/se-9-385-2018
Pareti, P., Konstantinidis, G., 2022. A Review of SHACL: From Data Validation to Schema Reasoning for RDF Graphs, in: Šimkus, M., Varzinczak, I. (Eds.), Reasoning Web. Declarative Artificial Intelligence, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 115–144. https://doi.org/10.1007/978-3-030-95481-9_6
Parquer, M.N., De Kemp, E.A., Brodaric, B., Hillier, M.J., 2024. Consistency-Checking 3D Geological Models. https://doi.org/10.5194/egusphere-2024-1326
Patel, D., Natali, M., Lidal, E.M., Parulek, J., Brazil, E.V., Viola, I., 2021. Modeling Terrains and Subsurface Geology, in: Patel, D. (Ed.), Interactive Data Processing and 3D Visualization of

the Solid Earth. Springer International Publishing, Cham, pp. 1–43. https://doi.org/10.1007/978-3-030-90716-7_1
Pérez, J., Arenas, M., Gutierrez, C., 2009. Semantics and complexity of SPARQL. ACM Transactions on Database Systems (TODS) 34, 1–45.
Perrin, M., Rainaud, J.-F., 2013. Shared earth modeling : knowledge driven solutions for building and managing subsurface 3D geological models. Technip Editions, Paris, 400pp.
Perrin, M., Zhu, B., Rainaud, J.-F., Schneider, S., 2005. Knowledge-driven applications for geological modeling. Journal of Petroleum Science and Engineering 47, 89–104. https://doi.org/10.1016/j.petrol.2004.11.010
Petcovic, H.L., Libarkin, J.C., Baker, K.M., 2009. An Empirical Methodology for Investigating Geocognition in the Field. Journal of Geoscience Education 57, 316–328. https://doi.org/10.5408/1.3544284
Pimenta, I., Kaschaka*, A., Cunha, F., 2005. Construction and validation of an integrated 3D geomodel for reservoir characterization – case study, offshore Angola, in: 9th International Congress of the Brazilian Geophysical Society & EXPOGEF, Salvador, Bahia, Brazil, 11-14 September 2005. Presented at the 9th International Congress of the Brazilian Geophysical Society & EXPOGEF, Salvador, Bahia, Brazil, 11-14 September 2005, Brazilian Geophysical Society, Salvador, Bahia, Brazil, pp. 1192–1197. https://doi.org/10.1190/sbgf2005-236
Pinto, H.S., Martins, J.P., 2004. Ontologies: How can They be Built? Know. Inf. Sys. 6, 441–464. https://doi.org/10.1007/s10115-003-0138-1
Pizzella, L., Alais, R., Lopez, S., Freulon, X., Rivoirard, J., 2022. Taking Better Advantage of Fold Axis Data to Characterize Anisotropy of Complex Folded Structures in the Implicit Modeling Framework. Mathematical Geosciences 54, 95–130. https://doi.org/10.1007/s11004-021-09950-0
Pochon, A., Gloaguen, E., Branquet, Y., Poujol, M., Ruffet, G., Boiron, M.-C., Boulvais, P., Gumiaux, C., Cagnard, F., Gouazou, F., Gapais, D., 2018. Variscan Sb-Au mineralization in Central Brittany (France): A new metallogenic model derived from the Le Semnon district. Ore Geology Reviews 97, 109–142. https://doi.org/10.1016/j.oregeorev.2018.04.016
Pochon, A., Poujol, M., Gloaguen, E., Branquet, Y., Cagnard, F., Gumiaux, C., Gapais, D., 2016. U-Pb LA-ICP-MS dating of apatite in mafic rocks: Evidence for a major magmatic event at the Devonian-Carboniferous boundary in the Armorican Massif (France). American Mineralogist 101, 2430–2442. https://doi.org/10.2138/am-2016-5736
Pollard, D.D., Fletcher, R.C., 2005. Fundamentals of structural geology. Cambridge university press, Cambridge (GB).
Ponciano, J.-J., 2019. Object detection in unstructured 3D data sets using explicit semantics (Theses). Université de Lyon.
Pylyshyn, Z.W., 1989. Computation and cognition: toward a foundation for cognitive science, 5. printing. ed, A Bradford book. MIT Press, Cambridge, Mass.
Qu, Y., Perrin, M., Torabi, A., Abel, M., Giese, M., 2023. GeoFault: A well-founded fault ontology for interoperability in geological modeling. arXiv preprint arXiv:2302.07059.
Ra, M., Yoo, D., No, S., Shin, J., Han, C., 2012. The mixed ontology building methodology using database information, in: Proceedings of the International MultiConference of Engineers and Computer Scientists.
Raab, T., Frodeman, R., 2002. What is it like to be a geologist? A phenomenology of geology and its epistemological implications. Philosophy & Geography 5, 69–81. https://doi.org/10.1080/10903770120116840
Ragan, D.M., 2009. Structural geology: an introduction to geometrical techniques, 4th ed. ed. Cambridge University Press, Cambridge ; New York.
Ramsay, J.G., 1974. Development of chevron folds. Geol Soc America Bull 85, 1741–1754. https://doi.org/10.1130/0016-7606(1974)85<1741:DOCF>2.0.CO;2
Ramsay, J.G., Huber, M.I., 2002. The techniques of modern structural geology. 2: Folds and fractures, 6. print. ed. Academic Pr, London.

Raskin, R.G., Pan, M.J., 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Computers and Geosciences 31, 1119–1125. https://doi.org/10.1016/j.cageo.2004.12.004

Rattanasawad, T., Saikaew, K.R., Buranarach, M., Supnithi, T., 2013. A review and comparison of rule languages and rule-based inference engines for the Semantic Web. Presented at the 2013 International Computer Science and Engineering Conference (ICSEC), IEEE, Nakorn Pathom, Thailand, pp. 1–6. https://doi.org/10.1109/ICSEC.2013.6694743

Raza, A., Gholami, R., Rezaee, R., Rasouli, V., Rabiei, M., 2019. Significant aspects of carbon capture and storage–A review. Petroleum 5, 335–340.

Renard, D., Lajaunie, C., Lopez, S., Allanic, C., Courrioux, G., Bourgine, B., Calcagno, P., 2019. La géostatistique au service de la modélisation géologique 3D. Annales des Mines - Responsabilité et environnement 94, 30–33. https://doi.org/10.3917/re1.094.0030

Requicha, Voelcker, 1982. Solid Modeling: A Historical Summary and Contemporary Assessment. IEEE Comput. Grap. Appl. 2, 9–24. https://doi.org/10.1109/MCG.1982.1674149

Rosenberg, G., 2001. An Artistic Perspective On The Continuity Of Space And The Origin Of Modern Geologic Thought. Earth Sciences History 20, 127–155. https://doi.org/10.17704/eshi.20.2.n845766556q63672

Rotevatn, A., Jackson, C.A.-L., 2014. 3D structure and evolution of folds during normal fault dip linkage. JGS 171, 821–829. https://doi.org/10.1144/jgs2014-045

Royse, K.R., 2010. Combining numerical and cognitive 3D modelling approaches in order to determine the structure of the Chalk in the London Basin. Computers and Geosciences 36, 500–511. https://doi.org/10.1016/j.cageo.2009.10.001

Rudolph, S., 2011. Foundations of Description Logics, in: Polleres, A., d'Amato, C., Arenas, M., Handschuh, S., Kroner, P., Ossowski, S., Patel-Schneider, P. (Eds.), Reasoning Web. Semantic Technologies for the Web of Data, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 76–136. https://doi.org/10.1007/978-3-642-23032-5_2

Samuel, A.L., 1959. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development 3, 210–229. https://doi.org/10.1147/rd.33.0210

Schweizer, D., Blum, P., Butscher, C., 2017. Uncertainty assessment in 3-D geological models of increasing complexity. Solid Earth 8, 515–530. https://doi.org/10.5194/se-8-515-2017

Segoni, S., Pappafico, G., Luti, T., Catani, F., 2020. Landslide susceptibility assessment in complex geological settings: Sensitivity to geological information and insights on its parameterization. Landslides 17, 2443–2453.

Sequeira, A.H., 2014. Conceptualization in Research. SSRN Journal. https://doi.org/10.2139/ssrn.2489284

Setianto, A., Triandini, T., 2015. Comparison of kriging and inverse distance weighted (idw) interpolation methods in lineament extraction and analysis. J. Appl. Geol. 5. https://doi.org/10.22146/jag.7204

Shipley, T.F., Tikoff, B., Ormand, C., Manduca, C., 2013. Structural geology practice and learning, from the perspective of cognitive science. Journal of Structural Geology 54, 72–84. https://doi.org/10.1016/j.jsg.2013.07.005

Shragge, J., Lumley, D., Bourget, J., Potter, T., Miyoshi, T., Witten, B., Giraud, J., Wilson, T., Iqbal, A., Emami Niri, M., Whitney, B., 2019. The Western Australia Modeling project — Part 2: Seismic validation. Interpretation 7, T793–T807. https://doi.org/10.1190/INT-2018-0218.1

Silva-Fragoso, A., Ferrari, L., Norini, G., Orozco-Esquivel, T., Corbo-Camargo, F., Bernal, J.P., Castro, C., Arrubarrena-Moreno, M., 2021. Geology and conceptual model of the Domuyo geothermal area, northern Patagonia, Argentina. Journal of Volcanology and Geothermal Research 420, 107396. https://doi.org/10.1016/j.jvolgeores.2021.107396

Simons, B., Boisvert, E., Brodaric, B., Cox, S., Duffy, T.R., Johnson, B.R., Laxton, J.L., Richard, S., 2006. GeoSciML: Enabling the Exchange of Geological Map Data, in: ASEG Extended Abstracts. Presented at the Conference: Australian Earth Sciences Convention, Melbourne, Australia, pp. 1–4. https://doi.org/10.1071/ASEG2006ab162

Sirin, E., Parsia, B., Grau, B.C., Kalyanpur, A., Katz, Y., 2007. Pellet: A practical OWL-DL reasoner. Journal of Web Semantics 5, 51–53. https://doi.org/10.1016/j.websem.2007.03.004

Smith, B., Malyuta, T., Rudnicki, R., Mandrick, W., Salmen, D., Morosoff, P., Duff, D.K., Schoening, J., Parent, K., 2013. Iao-Intel: An Ontology of Information Artifacts in the Intelligence Domain, in: Laskey, K.B., Emmons, I., Costa, P.C.G. (Eds.), Proceedings of the Eighth International Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS), CEUR, Vol. 1097. pp. 33–40.

Souche, L., Lepage, F., Iskenova, G., 2013. Volume Based Modeling - Automated Construction of Complex Structural Models. Presented at the 75th EAGE Conference and Exhibition incorporating SPE EUROPEC 2013, London, UK. https://doi.org/10.3997/2214-4609.20130037

Sousa, C., Soares, A.L., Pereira, C., Moniz, S., 2014. Establishing Conceptual Commitments in the Development of Ontologies through Competency Questions and Conceptual Graphs, in: Meersman, R., Panetto, H., Mishra, A., Valencia-García, R., Soares, A.L., Ciuciu, I., Ferri, F., Weichhart, G., Moser, T., Bezzi, M., Chan, H. (Eds.), On the Move to Meaningful Internet Systems: OTM 2014 Workshops, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 626–635. https://doi.org/10.1007/978-3-662-45550-0_64

Sowa, J.F., others, 1992. Semantic networks. Encyclopedia of artificial intelligence 2, 1493–1511.

Sprague, K.B., De Kemp, E.A., 2005. Interpretive tools for 3-D structural geological modelling part II: Surface design from sparse spatial data. GeoInformatica 9, 5–32.

Staab, S., Studer, R. (Eds.), 2009. Handbook on Ontologies. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-92673-3

Stead, D., Wolter, A., 2015. A critical review of rock slope failure mechanisms: the importance of structural geology. Journal of Structural Geology 74, 1–23.

Studer, R., Benjamins, V.R., Fensel, D., 1998. Knowledge engineering: Principles and methods. Data & Knowledge Engineering 25, 161–197. https://doi.org/10.1016/S0169-023X(97)00056-6

Tait, S.R., Jaupart, C., 1996. The production of chemically stratified and adcumulate plutonic igneous rocks. Mineral. mag. 60, 99–114. https://doi.org/10.1180/minmag.1996.060.398.07

Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Tarantola, A., Valette, B., others, 1982. Inverse problems= quest for information. Journal of geophysics 50, 159–170.

Terrington, R., Napier, B., Howard, A., Ford, J., Hatton, W., Oleschko, K., Cherkasov, S., Prieto, J.L.P., Argüelles, V.T., Salado, C.I.G., Miranda, A.G.C., Castro, S.A.Z., 2008. Why 3D? The Need for Solution Based Modeling in a National Geoscience Organization., in: AIP Conference Proceedings. Presented at the GIS IN GEOLOGY AND EARTH SCIENCES: 4th International Conference "In Vista of New Approaches for the Geoinformatics," AIP, Queretaro (Mexico), pp. 103–112. https://doi.org/10.1063/1.2937278

Thiele, S.T., Jessell, M.W., Lindsay, M., Ogarko, V., Wellmann, J.F., Pakyuz-Charrier, E., 2016a. The topology of geology 1: Topological analysis. Journal of Structural Geology 91, 27–38. https://doi.org/10.1016/j.jsg.2016.08.009

Thiele, S.T., Jessell, M.W., Lindsay, M., Wellmann, J.F., Pakyuz-Charrier, E., 2016b. The topology of geology 2: Topological uncertainty. Journal of Structural Geology 91, 74–87. https://doi.org/10.1016/j.jsg.2016.08.010

Tikoff, B., Blenkinsop, T., Kruckenberg, S.C., Morgan, S., Newman, J., Wojtal, S., 2013. A perspective on the emergence of modern structural geology: Celebrating the feedbacks between historical-based and process-based approaches, in: The Web of Geological Sciences: Advances, Impacts, and Interactions. Geological Society of America. https://doi.org/10.1130/2013.2500(03)

Torabi, A., Berg, S.S., 2011. Scaling of fault attributes: A review. Marine and Petroleum Geology 28, 1444–1460. https://doi.org/10.1016/j.marpetgeo.2011.04.003

Toueg, S., Babaoglu, Ö., 1984. On the Optimum Checkpoint Selection Problem. SIAM J. Comput. 13, 630–649. https://doi.org/10.1137/0213039

Tsang, C.-F., Neretnieks, I., Tsang, Y., 2015. Hydrologic issues associated with nuclear waste repositories. Water Resources Research 51, 6923–6972.

Turner, A.K., Gable, C.W., 2007. A review of geological modeling. Three-dimensional geologic mapping for groundwater applications 75–79.

Ulmer, K.J., Rodriguez-Marek, A., Green, R.A., 2021. Accounting for Epistemic Uncertainty in Site Effects in Probabilistic Seismic Hazard Analysis. Bulletin of the Seismological Society of America 111, 2005–2020. https://doi.org/10.1785/0120200343

Urvois, M., Grellet, S., Lorenz, H., Haener, R., Loiselet, C., Harrison, M., Krivic, M., Pedersen, C.B., Wiese, M.B., Baptie, A., Nayembil, M., Trench, J., Marsh, I., Cipolloni, C., D'Ambrogi, C., Congi, M.P., 2022. Integrating geological data in Europe to foster multidisciplinary research. Annals of Geophysics 65, DM319. https://doi.org/10.4401/ag-8817

Verhnet, Y., 2010. Craon. Carte géologique de la France à 1:50 000.

Verschuren, M., Butler, D.M., 2016. Effective glue between geoscience concepts, data, and modeling systems. Interpretation 4, T103–T121. https://doi.org/10.1190/int-2014-0259.1

Walton, D.N., 1990. What is Reasoning? What Is an Argument? The Journal of Philosophy 87, 399. https://doi.org/10.2307/2026735

Wang, C., Ma, X., Chen, J., 2018. Ontology-driven data integration and visualization for exploring regional geologic time and paleontological information. Computers and Geosciences 115, 12–19. https://doi.org/10.1016/j.cageo.2018.03.004

Wang, X., Gorlitsky, R., Almeida, J.S., 2005. From XML to RDF: how semantic web technologies will change the design of "omic" standards. Nat Biotechnol 23, 1099–1103. https://doi.org/10.1038/nbt1139

Wang, Z., Qu, H., Wu, Z., Wang, X., 2018. Geo3DML: A standard-based exchange format for 3D geological models. Computers and Geosciences 110, 54–64. https://doi.org/10.1016/j.cageo.2017.09.008

Wang, Z., Qu, H., Wu, Z., Yang, H., Du, Q., 2016. Formal representation of 3D structural geological models. Computers and Geosciences 90, 10–23. https://doi.org/10.1016/j.cageo.2016.02.007

Watson, D., 1999. The natural neighbor series manuals and source codes. Computers & Geosciences 25, 463–466. https://doi.org/10.1016/S0098-3004(98)00150-2

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 1st ed. Wiley. https://doi.org/10.1002/9780470517277

Wellmann, F., Caumon, G., 2018. 3-D Structural geological models: Concepts, methods, and uncertainties. Advances in Geophysics 59, 1–121. https://doi.org/10.1016/bs.agph.2018.09.001

Wellmann, J.F., De La Varga, M., Murdie, R.E., Gessner, K., Jessell, M., 2018. Uncertainty estimation for a geological model of the Sandstone greenstone belt, Western Australia – insights from integrated geological and geophysical inversion in a Bayesian inference framework. SP 453, 41–56. https://doi.org/10.1144/SP453.12

Wellmann, J.F., Finsterle, S., Croucher, A., 2014a. Integrating structural geological data into the inverse modelling framework of iTOUGH2. Computers & Geosciences 65, 95–109. https://doi.org/10.1016/j.cageo.2013.10.014

Wellmann, J.F., Lindsay, M., Poh, J., Jessell, M., 2014b. Validating 3-D Structural Models with Geological Knowledge for Improved Uncertainty Evaluations. Energy Procedia 59, 374–381. https://doi.org/10.1016/j.egypro.2014.10.391

Whitehead, B., Gahegan, M., 2012. Deep semantics in the geosciences: Semantic building blocks for a complete geoscience infrastructure. CEUR Workshop Proceedings 969, 74–85.

Wolf, K.H. (Ed.), 1976. Handbook of strata-bound and stratiform ore deposits. Elsevier Scientific Pub. Co, Amsterdam ; New York.

Wood, R., Curtis, A., 2004. Geological prior information and its applications to geoscientific problems. SP 239, 1–14. https://doi.org/10.1144/GSL.SP.2004.239.01.01

Wu, Q., Xu, H., 2003. An approach to computer modeling and visualization of geological faults in 3D. Computers & Geosciences 29, 503–509. https://doi.org/10.1016/S0098-3004(03)00018-9

Wu, Q., Xu, H., Zou, X., Lei, H., 2015. A 3D modeling approach to complex faults with multi-source data. Computers & Geosciences 77, 126–137. https://doi.org/10.1016/j.cageo.2014.10.008

Zanchi, A., Francesca, S., Stefano, Z., Simone, S., Graziano, G., 2009. 3D reconstruction of complex geological bodies: Examples from the Alps. Computers & Geosciences 35, 49–69. https://doi.org/10.1016/j.cageo.2007.09.003

Zhan, X., Liang, J., Lu, C., Hu, G., 2019. Semantic Description and Complete Computer Characterization of Structural Geological Models. Geoscientific Model Development Discussions 1–39. https://doi.org/10.5194/gmd-2018-305

Zhan, X., Lu, C., Hu, G., 2022. A Formal Representation of the Semantics of Structural Geological Models. Scientific Programming 2022. https://doi.org/10.1155/2022/5553774

Zhang, T., 2008. Incorporating Geological Conceptual Models and Interpretations into Reservoir Modeling Using Multiple-Point Geostatistics. Earth Science Frontiers 15, 26–35. https://doi.org/10.1016/S1872-5791(08)60016-0

Zhang, Y., Ling, C., 2018. A strategy to apply machine learning to small datasets in materials science. npj Comput Mater 4, 25. https://doi.org/10.1038/s41524-018-0081-z

Zhong, D., Wang, L., Bi, L., 2020. Extended Hermite Radial Basis Functions for Sparse Contours Interpolation. IEEE Access 8, 58752–58762. https://doi.org/10.1109/ACCESS.2020.2982802

Zhong, J., Aydina, A., McGuinness, D.L., 2009. Ontology of fractures. Journal of Structural Geology 31, 251–259. https://doi.org/10.1016/j.jsg.2009.01.008

Zimmerman, C., 2000. The development of scientific reasoning skills. Developmental review 20, 99–149.

Zlatanova, S., Rahman, A.A., Shi, W., 2004. Topological models and frameworks for 3D spatial objects. Computers & Geosciences 30, 419–428. https://doi.org/10.1016/j.cageo.2003.06.004

Zulauf, G., Zulauf, J., Maul, H., 2017. Quantification of the geometrical parameters of non-cylindrical folds. Journal of Structural Geology 100, 120–129. https://doi.org/10.1016/j.jsg.2017.06.001

Appendices

Appendix A

This appendix section details the proposed knowledge model (MOGI) including description of the different classes, their properties and relations. The full ontology, and the prototype algorithm with its source code can be found here*. The class Pointbased_Observation is a class for describing an observation that is generally represented by a point-like field observations. An instance of the class Pointbased_Observation has a representation an instance of the class Point through the relation has_Representation. Instances of the class Point have properties that allow localizing them in the space (coordinates). A Pointbased_Observation has properties that hold structural information (dip, dip_dir and polarity). An instance of the class Pointbased_Observation could be only explained by an instance of the Stratigraphic_Part class. This relation of explanation is enabled through the relation has_Possible_Explanation, and its inverse relation is_Possible_Explanation_Of like shown in Figure A-1.

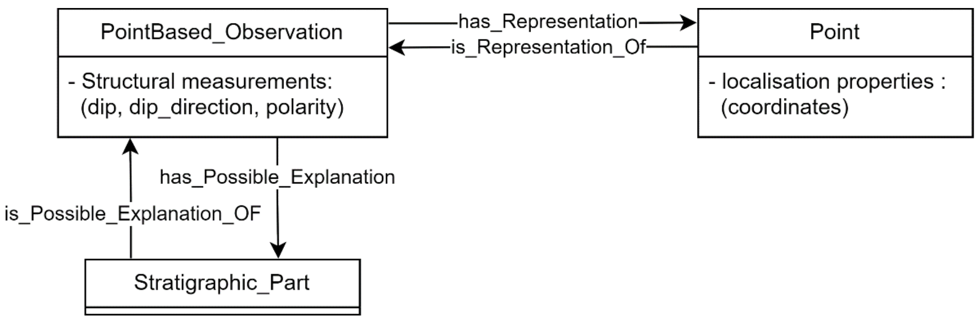


Figure A-1 An illustration detailing the relation and properties of the PointBased_Observation class.

The class Stratigraphic_Part, as shown in Figure A-2, stands for a portion of a geological surface (top or bottom of a unit) with defined size. This class has properties like deformation status (to state if it is dipping or not), polarity (younging direction), and size. A Stratigraphic_Surface has a property Stratigraphy to specify the stratigraphic formation in question. A Stratigraphic_Part could be explained by another Stratigraphic_Part. This relation of explanation is enabled through the relation has_Possible_Explanation. A Stratigraphic_Part could be represented by a surface, either planar or complex, and this is enabled through the has_Representation relation. If the representation is of the type Complex_Surface, the instance will be a collection of some planar surfaces. The instance of Complex_Surface is related to these surfaces by the relation has_Part and its inverse relation is_Part_Of. For the explanation to be valid, the instance of the stratigraphic_part must cover the selected features (instances of PointBased_Observation or Stratigraphic_Part) within a margin of tolerance. In addition, the explaining instance must be consistent with the orientation of the selected features.

* https://github.com/BRGM/MOGI

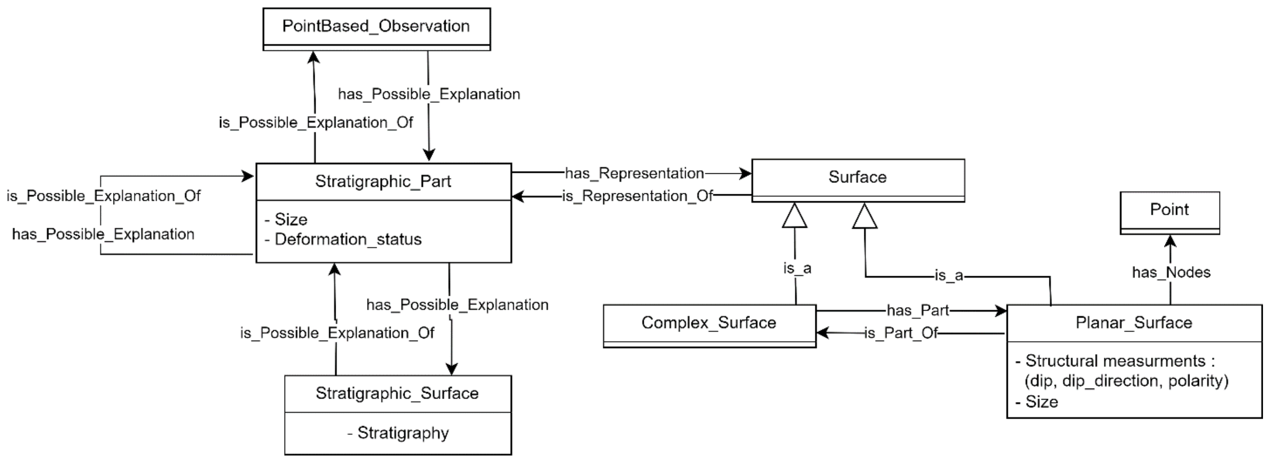


Figure A-2 An illustration detailing the relation and properties of the Stratigraphic_Part class

To explain when a geological layer is deformed, the concept fold is used. For the simplification, the only deformation described in the ontological model is folding through the class Fold. Also, for simplification, chevron folding is the only type of folding used through the class Chevron_Fold (Figure A-3). Chevron folding is defined as trend "repeated set of v-shaped beds with straight limbs and sharp hinges" (Ramsay, 1974). Folds are characterized in this proof of concept by two structural parameters: an axial surface and a fold axis. Thus, classes Axial_Surface and Fold_Axis are used. Their geometric representation is a planar surface and a vector respectively. An instance of the class Chevron_Fold is related to its structural parameters by the relations has_Axial_Surface and has_Fold_Axis. A Chevron_Fold is also characterized by its two limbs, which are represented by two instances of the Class Fold_Limb. The two limbs are of the type stratigraphic part, with each having its own representation.

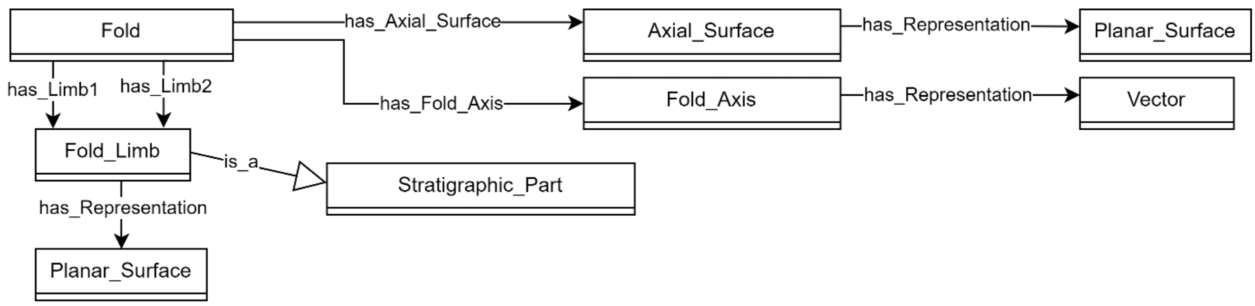


Figure A-3 An illustration detailing the relation and properties of the Fold class

To specify which limb is considered as first or second, the relations has_Limb1 and has_Limb2 are used. Such distinction is important to characterize the direction of fold axis and gives a proper orientation for the fold. For a fold to be accepted, it must have valid structural parameters, its limbs must have valid polarity and, in the space where the fold is defined, it must be free of other stratigraphic parts that are not explained by the fold. The existence of multiple folds is explained by a fold train object that is an instance of the class Fold_Train.

An anomaly is an unexpected or invalid situation that could be detected between inputs during the process of interpretation. The anomaly class has eighteen subclasses, each specifying a distinct anomaly that could be encountered during the interpretation process as depicted in Figure A-4. Each anomaly can be related to specific entities (an object, a part of an object, or a representation) by the relation is_Related_To. Each anomaly can be dealt with by one or several actions (new interpretation, delete, upgrade). These actions are not implemented in the ontology, but only in the interpretation block (IM).

Figure A-4 The 18 anomaly subclasses. These denote anomalous situations that could be encountered during the interpretation process and the representation of the model

Appendix B

This appendix section includes additional figures (from A-5 to A13) from the applications used to address the proposed use cases. It primarily contains snapshots from the Protégé software, showing the POKIMON implementation of the main classes, and the ontological representation of key elements in the use cases, along with SPARQL queries for each use case to demonstrate the information retrieval capabilities of the proposed POKIMON model. The full ontology, and the examples, SPARQL queries can be found here*.

*https://github.com/Imadont/POKIMON

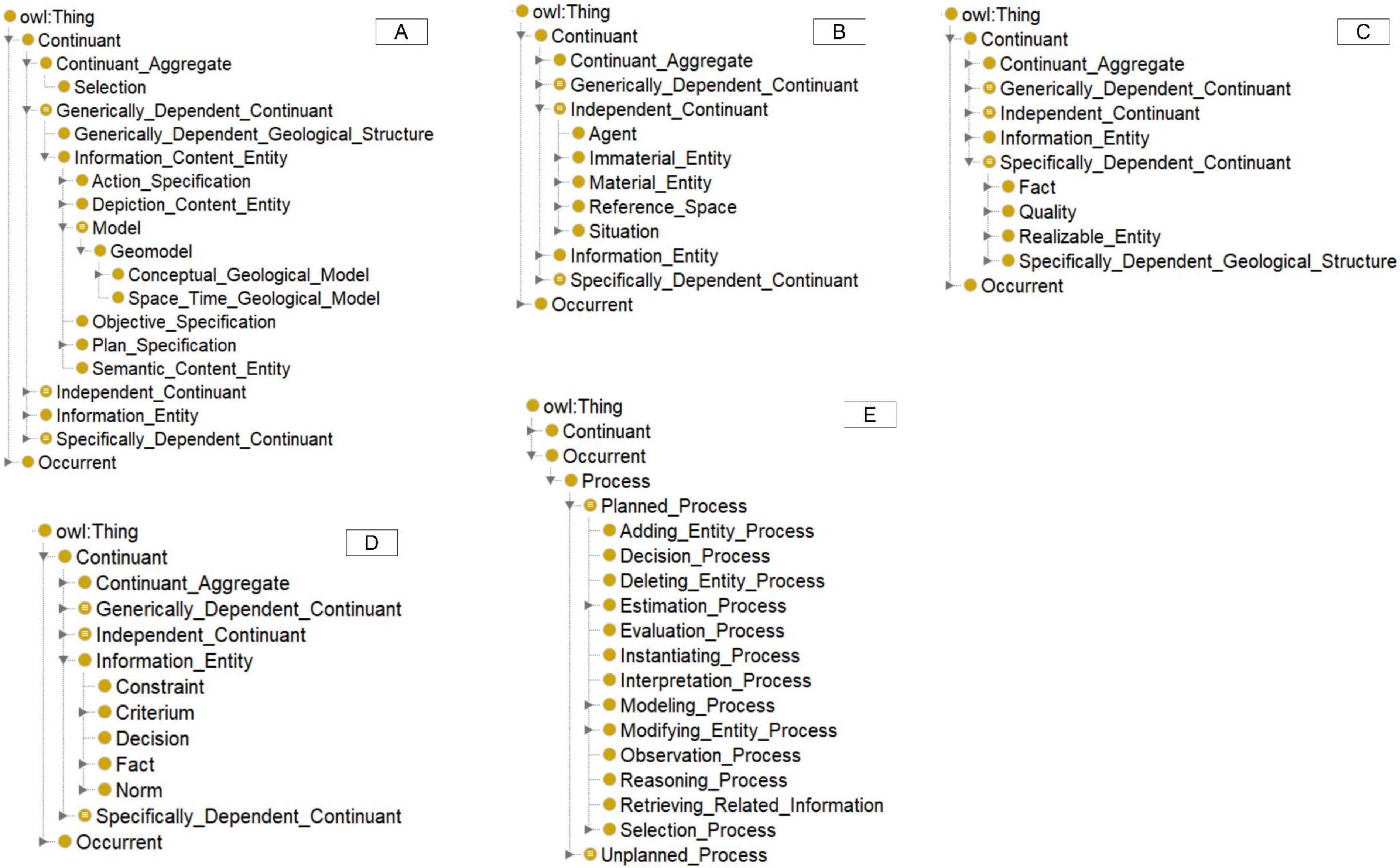


Figure B-1 Plate showing the main classes of POKIMON. (A) shows the subclasses classes of the Generically Dependent Continuant class, with focus on the Geomodel class. (B) shows the subclasses of the Independent Continuant class. (C) shows the subclasses of the Specifically Dependent Continuant class. (D) shows the subclasses of the information entity class. (E) shows the subclasses of the occurrent with focus on the subclasses of the planned process class

A

```
SELECT DISTINCT ?geomodels  ?sema_cont    ?depic_cont
WHERE  {
          {?geomodels rdf:type pok:Geomodel} UNION
                  {?geomodels rdf:type/rdfs:subClassOf* pok:Geomodel}
         ?geomodels pok:has_Depiction_Component ?depic_cont ;
                  pok:has_Semantic_Component ?sema_cont
```

geomodels	sema_cont	depic_cont
conceptual_geomodel_folded_surfs	semantic_comp_conceptual	cartoonish_representation_for_conc_geom
space_time_geomodel_folded_surfs	semantic_comp_space_time	representation_3D_of_folded_boundaries
space_time_geomodel_folded_surfs	semantic_comp_space_time	textual_description_of_folded_boundaries

B

```
SELECT DISTINCT  ?sema_cont ?geol_ent
WHERE  {
          ?sema_cont  rdf:type pok:Semantic_Content_Entity;
                  pok:has_Semantic_Component ?geol_ent

}
```

sema_cont	geol_ent
semantic_comp_conceptual	boundary_1
semantic_comp_conceptual	boundary_3
semantic_comp_conceptual	boundary_2
semantic_comp_space_time	part_of_folded_boundary_1
semantic_comp_space_time	part_of_folded_boundary_3
semantic_comp_space_time	part_of_folded_boundary_2

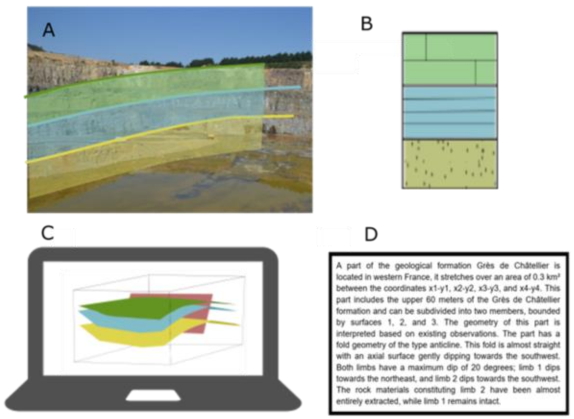


Figure B-2 Plate showing the SPARQL queries for the use case illustrated in the lower right corner (refer to use-case 1 for more details). The ontological representation can be found in Figure 3-17. Panel (A) focuses on the geomodels, highlighting their semantic and depiction components. Panel (B) presents the geological entities within the two semantic components

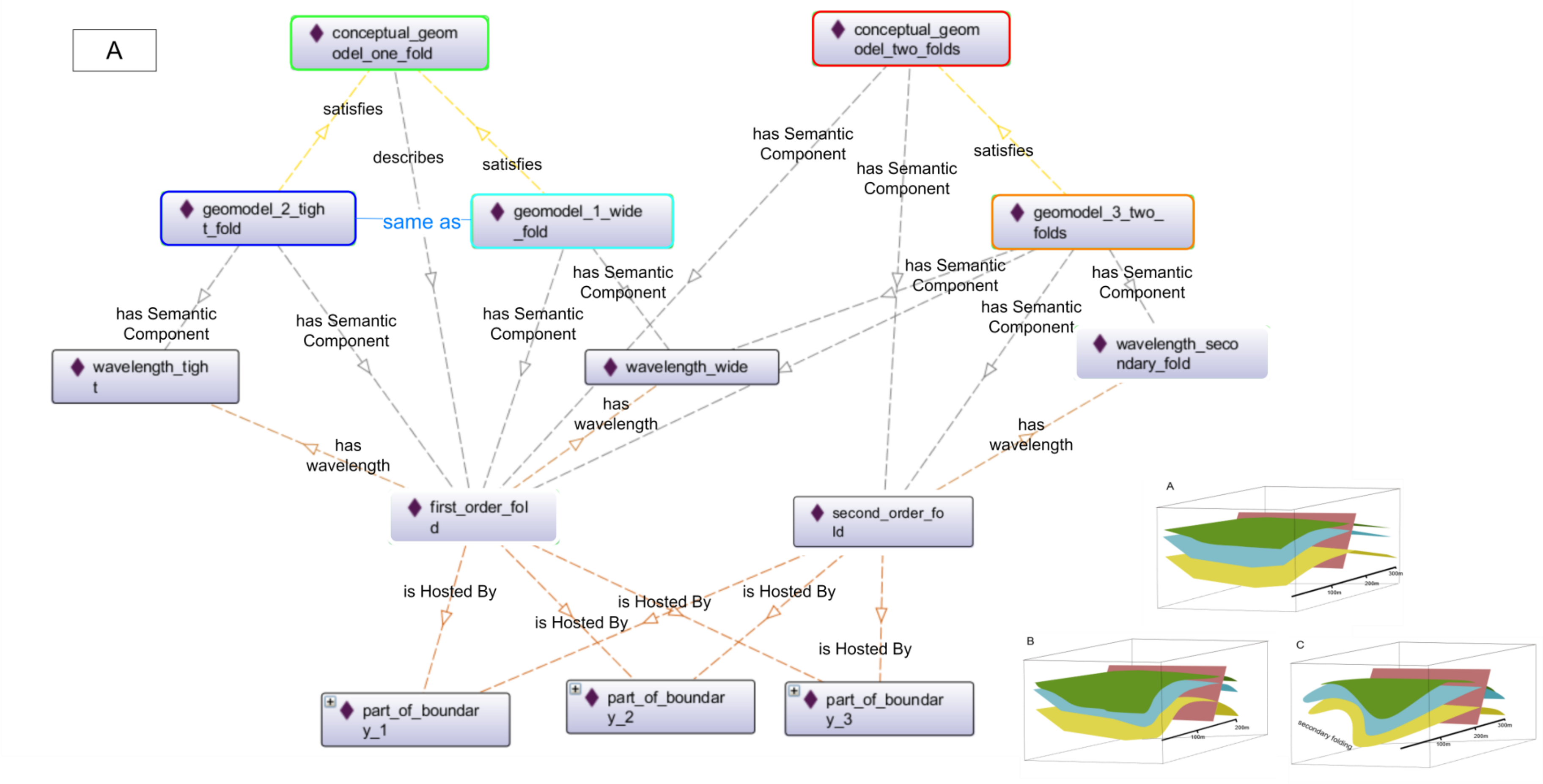


Figure B-3 A-7 Plate (A) showing the ontological representation for the use case illustrated in the lower right corner (refer to use-case 2 for more details), SPARQL queries are provided in the next page (B, C). the modeling of the original geomodel is highlighted by the green frame. This conceptual geomodel is satisfied by a space time geomodel (geomodel 1 wide fold in navy blue). When creating an updated of this space time geomodel with modified properties (wavelength quality of the associated fold), the new created space time geomodel (geomodel 2 tight fold, cyan blue) it is considered as the same (same as relation) of the first space time geomodel. When another geomodel is proposed with more updates (additional fold 'secondary fold), this space time geomodel (geomodel 3 two folds, orange frame) is considered a different geomodel as it satisfies a different conceptual geomodel (highlighted in red)

B

```
SELECT DISTINCT ?space_time_geom  ?concept_geom ?geol_ent
WHERE  {
        ?space_time_geom rdf:type pok:Space_Time_Geological_Model;
                pok:satisfies ?concept_geom;
                pok:has_Semantic_Component ?geol_ent .
}
```

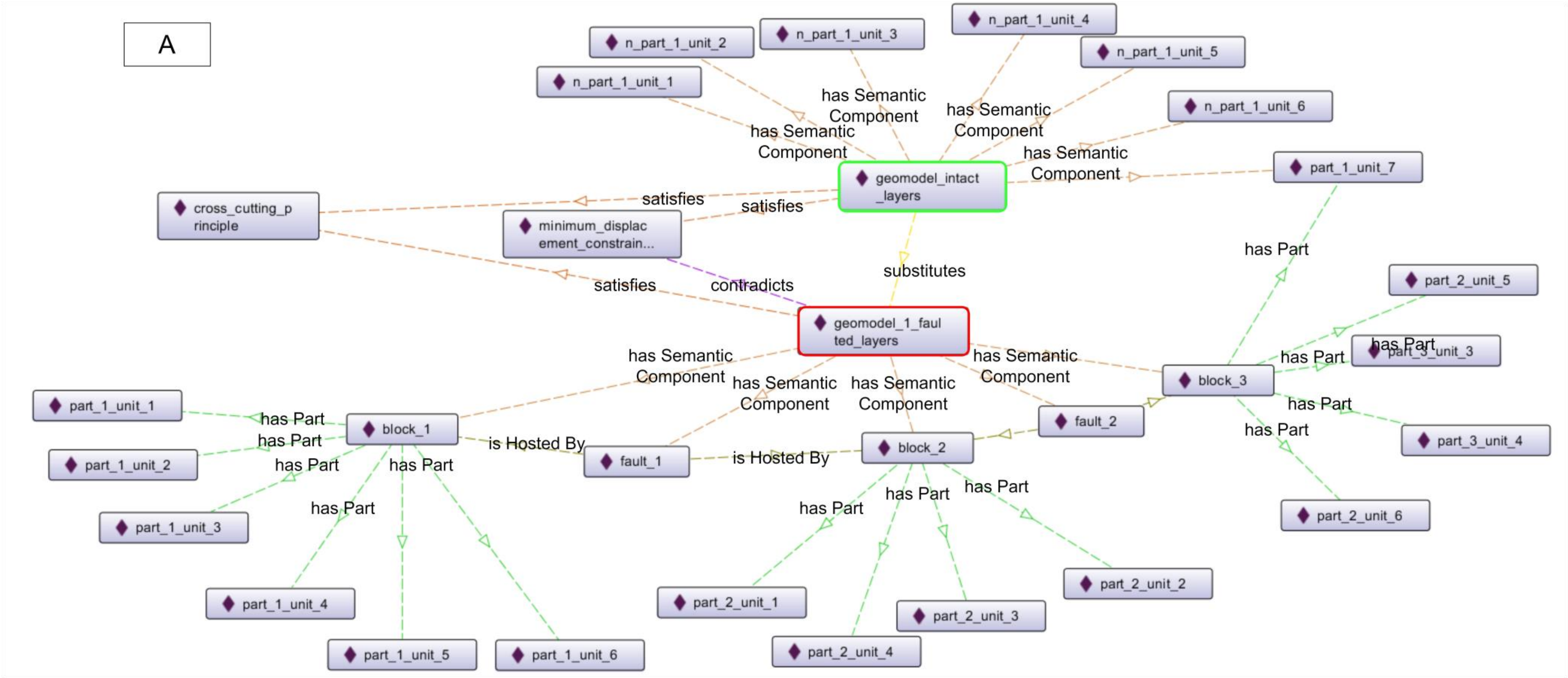
space_time_geom	concept_geom	geol_ent
geomodel_1_wide_fold	conceptual_geomodel_one_fold	wavelength_wide
geomodel_1_wide_fold	conceptual_geomodel_one_fold	first_order_fold
geomodel_2_tight_fold	conceptual_geomodel_one_fold	wavelength_tight
geomodel_2_tight_fold	conceptual_geomodel_one_fold	first_order_fold
geomodel_3_two_folds	conceptual_geomodel_two_folds	wavelength_secondary_fold
geomodel_3_two_folds	conceptual_geomodel_two_folds	seond_order_fold
geomodel_3_two_folds	conceptual_geomodel_two_folds	first_order_fold
geomodel_3_two_folds	conceptual_geomodel_two_folds	wavelength_wide

C

```
SELECT DISTINCT  ?concept_geom ?geol_ent
WHERE  {
        ?concept_geom rdf:type pok:Conceptual_Geological_Model;
                pok:has_Semantic_Component ?geol_ent .
}
ORDER BY ?space_time_geom
```

concept_geom	geol_ent
conceptual_geomodel_one_fold	first_order_fold
conceptual_geomodel_two_folds	seond_order_fold
conceptual_geomodel_two_folds	first_order_fold

Part 2 Figure B-3 A-7 SPARQL queries for the use case about the geomodel integrity detailing (A) each space-time geomodel, the conceptual geomodel it satisfies, and (C) the entities described by all these geomodels



B

```
SELECT ?geomodels  ?relation ?relatedModel
WHERE {
  ?geomodels rdf:type pok:Space_Time_Geological_Model;
          ?relation ?relatedModel .
  ?relatedModel rdf:type pok:Space_Time_Geological_Model.
}
```

geomodels	relation	relatedModel
geomodel_intact_layers	substitutes	geomodel_1_faulted_layers

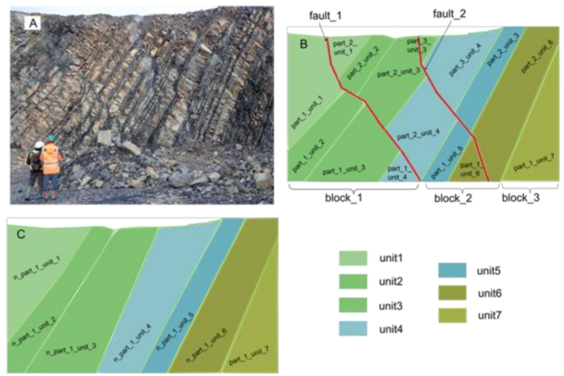
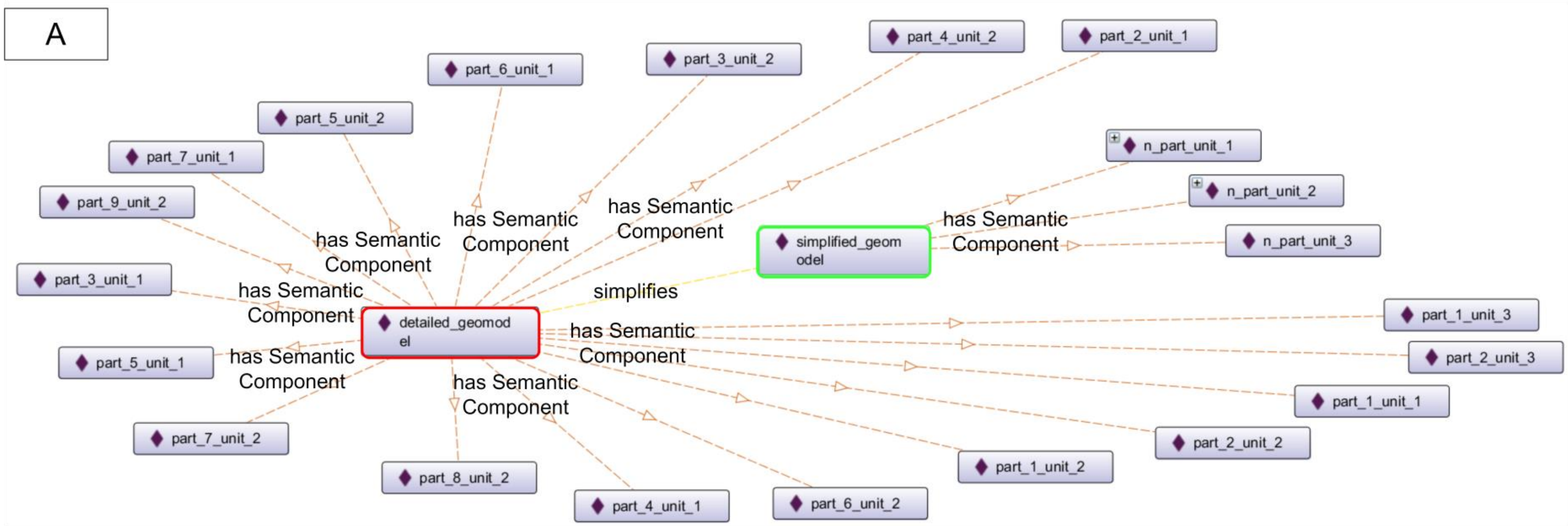


Figure B-4 Schematic plate showing the modeling of the key elements of the use case illustrated in the lower right corner (refer to use-case 3 for more details). (A) in green a space time geomodel with non-faulted units that substitutes the space time geomodel in red where units are faulted but where the displacement value is below (contradicting) the constraints set for the use case. (B) shows a SPARQL query for which geomodel substitutes the other one



B

```
SELECT ?geomodels  ?relation ?relatedModel
WHERE {
  ?geomodels rdf:type pok:Space_Time_Geological_Model;
           ?relation ?relatedModel .
  ?relatedModel rdf:type pok:Space_Time_Geological_Model.
}
```

geomodels	relation	relatedModel
simplified_geomodel	simplifies	detailed_geomodel

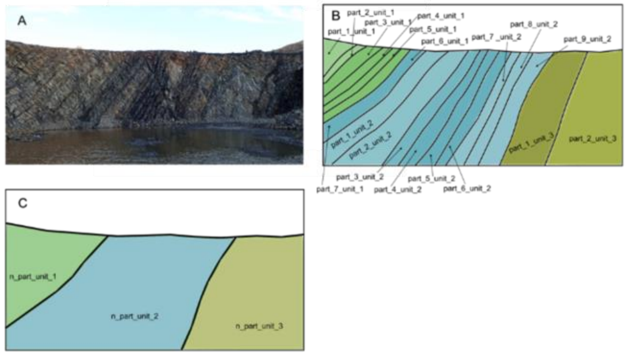


Figure B-5 Schematic plate showing the modeling of the key elements of the use case illustrated in the lower right corner (refer to use-case 3 for more details). (A) in green a space time geomodel with 3 units that simplifies the more detailed space-time geomodel in red, which includes descriptions of inner beds. Panel (B) displays a SPARQL query indicating which geomodel serves as a simplification of the other

A

```
SELECT ?interpretation_processes ?inputs
        WHERE { ?interpretation_processes a pok:Interpretation_Process;
                                    pok:has_Input ?inputs
}
```

interpretation_processes	inputs
process_intepretation_fold1	gmd_10
process_intepretation_fold1	gmd_7
process_intepretation_fold1	gmd_6
process_intepretation_fold1	gmd_5
process_intepretation_fold1	gmd_4
process_intepretation_fold1	gmd_9
process_intepretation_fold1	gmd_8
process_intepretation_fold1	gmd_3
process_intepretation_fold1	gmd_2
process_intepretation_fold1	gmd_1

B

```
SELECT ?interpretation_processes ?outputs
        WHERE { ?interpretation_processes a pok:Interpretation_Process;
                                    pok:has_Output ?outputs
}
```

interpretation_processes	outputs
process_intepretation_fold1	geomodel_fold1

C

```
SELECT ?interpretation_processes ?algorithms
        WHERE { ?interpretation_processes a pok:Interpretation_Process;
                                    pok:is_Prescribed_By ?algorithms
}
```

interpretation_processes	algorithms
process_intepretation_fold1	algo_interpretation

D

```
SELECT ?interpretation_processes ?parts
        WHERE { ?interpretation_processes a pok:Interpretation_Process;
                                    pok:has_Part ?parts
}
```

interpretation_processes	parts
process_intepretation_fold1	process_geomodel_fold1_realization
process_intepretation_fold1	process_evaluation_geomodel_fold1
process_intepretation_fold1	process_situ_select_1
process_intepretation_fold1	process_conceptual_model_fold1_proposition
process_intepretation_fold1	process_decision_geomodel_fold1

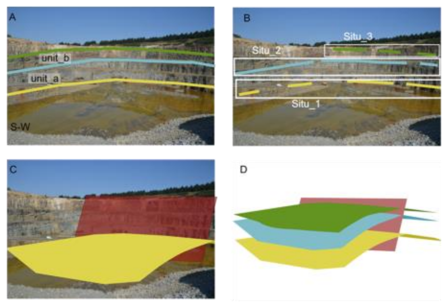


Figure B-6 Plate showing the SPARQL queries for the use case illustrated in the lower right corner (refer to use-case 4 for more details) depicting the modeling an iteration of an interpretation process. The ontological representation can be found in Figure 3-17. Panel (A) focuses on the inputs of the interpretation process, (B) on outputs, (C) on the algorithm that prescribes the process, and (D) on the different parts (subprocess) run within the general interpretation process

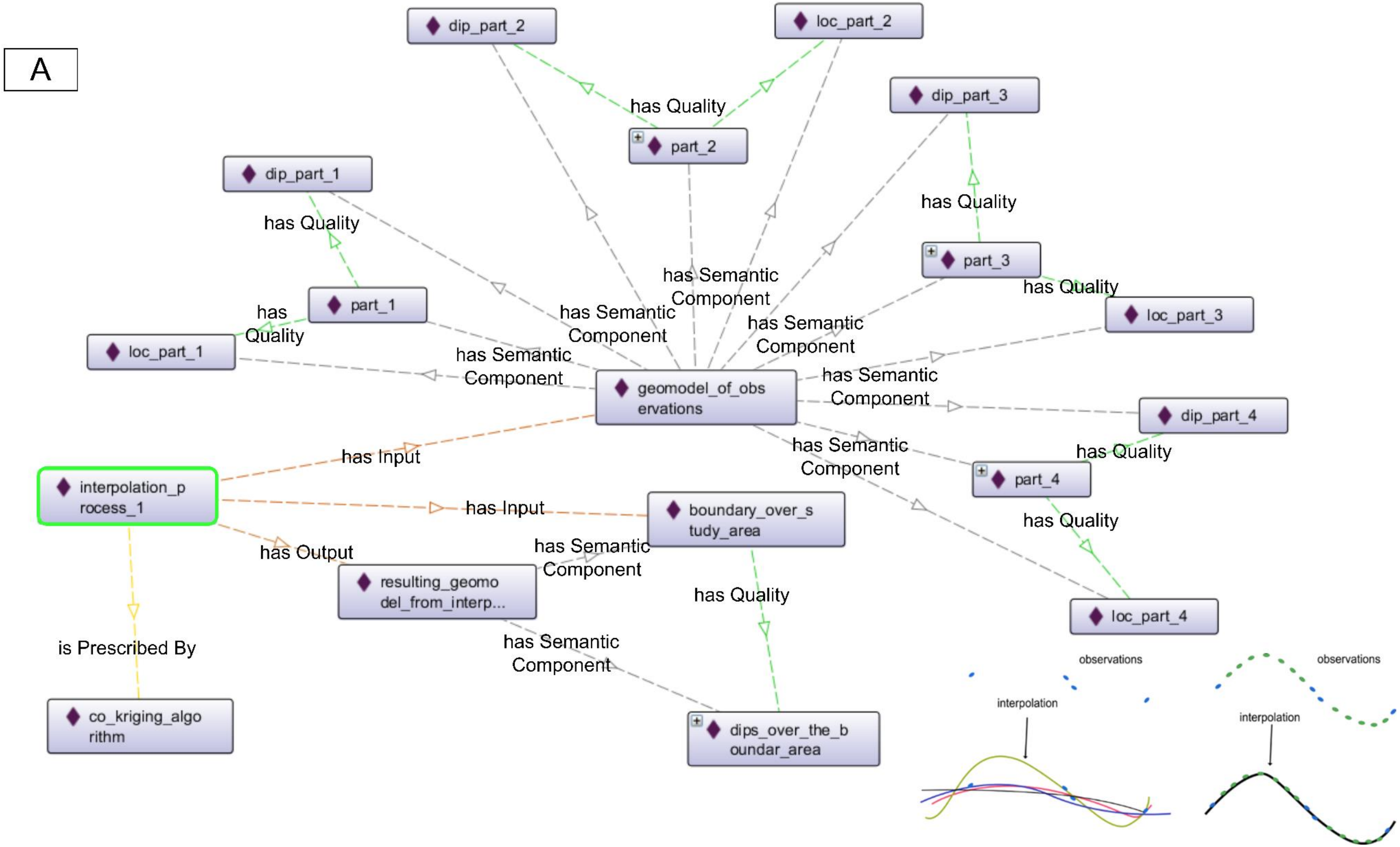


Figure B-7 Plate showing the SPARQL queries for the use case illustrated in the lower right corner (refer to use-case 5 for more details). The interpolation process used to generate the green surface is highlighted in green (A) with its inputs, outputs and the algorithm that describes it

B

```
SELECT DISTINCT ?geomodels  ?geol_ent
WHERE  {
            {?geomodels rdf:type pok:Geomodel} UNION
                    {?geomodels rdf:type/rdfs:subClassOf* pok:Geomodel}
        ?geomodels pok:has_Semantic_Component ?geol_ent .
}
ORDER BY ?geomodels ?geol_ent
```

geomodels	geol_ent
geomodel_of_observations	dip_part_1
geomodel_of_observations	dip_part_2
geomodel_of_observations	dip_part_3
geomodel_of_observations	dip_part_4
geomodel_of_observations	loc_part_1
geomodel_of_observations	loc_part_2
geomodel_of_observations	loc_part_3
geomodel_of_observations	loc_part_4
geomodel_of_observations	part_1
geomodel_of_observations	part_2
geomodel_of_observations	part_3
geomodel_of_observations	part_4
resulting_geomodel_from_interpolation	boundary_over_study_area
resulting_geomodel_from_interpolation	dips_over_the_boundar_area

C

```
SELECT DISTINCT ?interpolation_pross  ?algo  ?inputs
WHERE {
          ?interpolation_pross rdf:type pok:Interpolation_Process;
                                 pok:is_Prescribed_By ?algo;
                                 pok:has_Input ?inputs;

}

ORDER BY ?inputs
```

interpolation_pross	algo	inputs
interpolation_process_1	co_kriging_algorithm	boundary_over_study_area
interpolation_process_1	co_kriging_algorithm	geomodel_of_observations

interpolation_pross	algo	outputs
interpolation_process_1	co_kriging_algorithm	resulting_geomodel_from_interpolation

Part 2 Figure B-7 Plate showing the SPARQL queries for the previous use-case. Panel (B) presents the geological entities of the initial geomodel (geomodel of observations) that is considered as the inputs for the interpolation process. (C) shows the querying of this process (inputs, outputs, and prescribing algorithm)

Imadeddine LAOUICI

Formalisation des connaissances géologiques et automatisation du processus d'interprétation structural pour la construction d'architectures géologiques 3D du sous-sol

Les systèmes actuels utilisés pour construire des modèles géologiques structuraux 3D offrent des solutions performantes pour les représentations géométriques, mais reposent souvent sur les interprétations des experts et le formalisme mathématique en place. L'un des principaux verrous scientifiques réside dans l'incapacité de ces systèmes à s'impliquer dans le processus d'interprétation, en raison de l'utilisation tacite des connaissances par les experts. Cela entraîne des difficultés dans le partage et l'intégration des connaissances au sein des systèmes, ainsi qu'une complexité accrue pour reproduire et communiquer les modèles. Ce travail de recherche explore une approche innovante de modélisation, axée sur la notion d'interprétation et l'utilisation explicite des connaissances. Deux contributions majeures sont proposées : un formalisme pour l'interprétation et un cadre ontologique pour formaliser les connaissances mobilisées par les experts durant le processus de modélisation.

Mots clés : Modélisation géologique structurale 3D, Formalisme d'interprétation, Modélisation conceptuelle, Formalisation des connaissances, Ontologie, Incertitudes épistémiques.

Geological knowledge formalization and automation of the structural interpretation process for building 3D architectures of the sub-surface

Current systems used to build 3D structural geological models offer essential solutions for geometric representations, but often rely on expert interpretations and the existing mathematical formalism. One of the scientific challenges is the inability of these systems to be deeply involved in the interpretation process due to the tacit use of knowledge by experts. This leads to problems of sharing and integrating knowledge in systems, as well as issues in reproducing and communicating models. This research explores an innovative modeling approach, centered on the notion of interpretation and the explicit use of knowledge. In this context, two major contributions are proposed: an interpretation formalism supported by a prototype algorithm, and an ontological framework to formalize the knowledge used by experts during the modeling process.

Keywords : 3D Structural geological modelling Interpretation formalism, Conceptual modelling, Knowledge formalization, Ontology, Epistemic uncertainties.

Institut des Sciences de la Terre d'Orléans

1A rue de la Férollerie

45160 Orléans

Géosciences pour une Terre durable