



HAL
open science

Raisonner en logique modale sur l'incertitude liée aux données : application en archéologie

Elisabeth Gavignet, Eric Leclercq, Nadine Cullot, Marinette Savonnet

► To cite this version:

Elisabeth Gavignet, Eric Leclercq, Nadine Cullot, Marinette Savonnet. Raisonner en logique modale sur l'incertitude liée aux données : application en archéologie. *Revue Internationale de Géomatique*, 2016, Modéliser les dynamiques spatiales, 26 (4), pp.467-490. 10.3166/riig.2016.00007 . hal-01467036

HAL Id: hal-01467036

<https://hal.science/hal-01467036>

Submitted on 14 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Raisonner en logique modale sur l'incertitude liée aux données : application en archéologie

Élisabeth Gavignet, Éric Leclercq, Nadine Cullot,
Marinette Savonnet

Laboratoire LE2I - UMR6306 - CNRS - ENSAM
Univ. Bourgogne Franche-Comté
9, Avenue Alain Savary
F-21078 Dijon - France
Prenom.Nom@u-bourgogne.fr

RÉSUMÉ. La prise en compte de données imparfaites est un enjeu essentiel pour les systèmes d'information archéologiques, notamment parce que la constitution d'un corpus est un travail intrinsèquement collaboratif et interdisciplinaire. L'imperfection des données collectées est en partie liée à la fiabilité des sources qui contribuent à les enrichir. Notre contribution porte sur l'utilisation de plusieurs logiques modales pour modéliser et tester les effets de la prise en compte de données incertaines, mais aussi pour faciliter la vérification de la qualité du corpus dans une plateforme ouverte et extensible permettant de gérer les données archéologiques. Pour la partie applicative, le raisonneur utilisé, basé sur la logique du premier ordre, fournit aux archéologues les moyens de valider ou non leurs hypothèses. Plusieurs raisonnements avec différentes modalités sont proposés autour d'un exemple simple qui rassemble les types d'incertitudes habituellement rencontrés.

ABSTRACT. The consideration of imperfect data is a key issue for archaeological information systems, in particular because the elaboration of a corpus is intrinsically a collaborative and interdisciplinary task. The imperfect property of collected data is partly related to the reliability of the sources contributing to enrich them. Our contribution brings the use of several modal logics to model and verify the effects of the consideration of uncertain data, but also to check the quality of the corpus in an open and extensible platform allowing the management of archaeological data. From the application point of view the reasoner used, based on the first order logic, provides the archaeologists with tools to validate or not their hypotheses. Some case studies of reasoning using different modalities are proposed around a simple example which gathers the kinds of uncertainty usually encountered.

MOTS-CLÉS : raisonnement, logiques modales, incertitudes, annotations sémantiques, ontologies

KEYWORDS : reasoning, modal logics, uncertainty, semantic annotations, ontology

DOI:10.3166/RIG.26.1-22 © 2016 Lavoisier

1. Introduction

Les données archéologiques proviennent de sources multiples (fouilles, données cartographiques anciennes et actuelles, sources historiques, etc.). Leur collecte a recours à différentes techniques (de datation, de caractérisation des matériaux, etc.). Elle est réalisée sur une longue durée par de nombreux spécialistes. Le nombre important des acteurs impliqués et l'échelonnement dans le temps posent néanmoins le problème de la qualité des informations recueillies et notamment de leur cohérence les unes avec les autres.

L'évolution des données et notamment l'ajout de nouvelles connaissances dans un ensemble déjà existant nécessite le contrôle de leur cohérence soit de façon globale, soit de façon contextuelle selon les différentes sources d'information ou les interprétations des différents acteurs. Pour répondre à cette préoccupation, de nombreux travaux, basés sur une représentation des connaissances à l'aide de modèles logiques offrant des formalismes de modélisation et de raisonnement, ont été menés. Par exemple dans le cadre de la représentation des connaissances en logique classique, les approches basées sur les stratégies de révision (Doukari, Jeansoulin, 2007) permettent de déterminer les sous-ensembles minimaux de connaissances à retirer pour rétablir la cohérence des informations lors de l'ajout de nouvelles connaissances. Ces approches ont été exploitées dans plusieurs projets dont le projet européen VENUS (Virtual Exploration of Underwater Site) (Hué *et al.*, 2011). Cependant les données collectées évoluant, provenant de sources pas toujours fiables et étant susceptibles de présenter différents degrés d'imperfection (imprécision, incertitude, incohérence), d'autres approches basées sur des logiques non classiques s'attachent à modéliser toutes les connaissances et leurs imperfections.

Cet article se propose d'expliciter différentes possibilités pour représenter l'incertitude au sein d'un système évolutif, s'appuyant sur des modèles de représentation et d'interprétation de données incertaines, et offrant des outils de raisonnement et de détection des incohérences (Leclercq, Savonnet, 2012).

De nombreux travaux ont utilisé les logiques modales pour modéliser les connaissances et proposer une représentation des données incertaines selon différents contextes (sources, acteurs, etc.). Ils ont été appliqués dans des domaines aussi variés que la sécurité des informations personnelles (Piolle, Demazeau, 2008) en logique déontique, les incidences géométriques (Balbiani *et al.*, 1997) avec des relations temporelles, la réglementation dans des systèmes multi-agents (Garion *et al.*, 2010) reformulée dans un cadre modal du premier ordre, la définition de régions dans l'espace et de raisonnement spatial (Bentham, Bezhanishvili, 2007 ; Bennett, 1996). Cependant, ils n'implémentent pas, à notre connaissance, de moteurs d'inférences permettant de raisonner sur ces données. Notre approche vise à proposer un début de solution pour répondre à deux problématiques essentielles : la première est la possibilité de faire évoluer les modèles de données sans remettre en cause les applications qui les utilisent et la seconde est la préservation de la qualité des données et des connaissances même incertaines. Pour cela nous avons proposé une extension

du modèle d'annotations aux données incertaines en nous appuyant sur les travaux précédemment menés pour la plateforme WikiBridge (Leclercq, Savonnet, 2012).

La suite de l'article est organisée en 4 sections. La section 2 présente un état de l'art sur les typologies de l'incertitude à partir de classifications reconnues, montre l'intérêt des logiques modales pour modéliser certaines formes de connaissances incertaines et s'intéresse aux méthodes de raisonnement sur ces connaissances. La section 3 décrit l'architecture de la plateforme WikiBridge, présente une synthèse du modèle d'annotation (Leclercq, Savonnet, 2012) et précise le rôle des ontologies dans la plateforme. La section 4 développe l'utilisation des logiques modales pour des cas distincts de raisonnement : la gestion de la localisation, la datation et l'interprétation multi-acteurs (ou agents). La section 5 dresse le bilan de nos travaux et présente des perspectives plus larges d'usage des logiques modales comme base pour la modélisation, l'interrogation et le raisonnement sur des données incertaines dans les systèmes d'information scientifiques.

2. État de l'art

Dans cette partie nous décrivons, de manière synthétique, les classifications de l'incertitude les plus connues ainsi que les outils de modélisation qui peuvent être mis en œuvre pour représenter les données incertaines.

2.1. Typologies des incertitudes

L'usage dans la littérature de nombreux synonymes est caractéristique de la difficulté à définir ce que l'incertitude recouvre. Les terminologies pour exprimer l'incertitude sont vagues et aussi nombreuses que les disciplines qui l'étudient. Les principales définitions de l'incertitude que l'on retrouve dans les dictionnaires tournent autour de deux axes : (1) état de quelqu'un qui ne dispose pas de la connaissance nécessaire pour prendre une décision, (2) propriété physique de l'information consécutive aux limites de perception de la réalité. Ces deux définitions ne sont bien évidemment pas complètement indépendantes puisque une information incertaine peut induire de l'incertitude dans l'esprit d'une personne. Plusieurs classifications de l'incertitude ont été proposées dans la littérature, il n'est pas dans notre intention d'en donner un inventaire complet mais d'en citer quelques unes qui permettent de mieux cibler la place de notre contribution.

(Thomson *et al.*, 2005) proposent une répartition de neuf critères d'incertitude dans trois catégories : qualité de l'information, cohérence de l'information et objectivité de l'information (tableau 1). Ces catégories mettent en avant trois aspects de la connaissance : l'information prise indépendamment de son contexte doit répondre à certains critères pour être considérée de qualité, l'ensemble des informations sur un sujet donné doivent être cohérentes entre elles et enfin, l'importance de la source qui se doit d'être la plus objective possible.

Tableau 1. Synthèse de la typologie de l'incertitude de Thomson et al.

Qualité	Cohérence	Objectivité
exactitude de la collecte précision de la mesure exhaustivité des informations	concordance entre données transformation lacune temporelle	fiabilité des sources subjectivité indépendance de la source

(Skeels *et al.*, 2010) décrivent une hiérarchisation de l'incertitude en plusieurs niveaux (trois niveaux hiérarchiques et deux niveaux transversaux). Le niveau le plus faible d'incertitude est relatif à la précision de la mesure, le niveau moyen fait référence à la complétude des informations (celles dont on dispose, celles dont on ne dispose pas mais dont on connaît l'existence et celles dont on ignore l'existence) et le niveau le plus élevé correspond aux données produites par des inférences (supposition, extrapolation). Quant aux niveaux transversaux, ils concernent les notions de conflit et de crédibilité, lesquels peuvent se produire sur les trois niveaux hiérarchiques.

D'autres travaux comme (Gershon, 1998) et (Smets, 1997) élaborent une classification non pas de l'incertitude mais de l'imperfection. Gershon considère l'incertitude comme l'une des six causes de l'imperfection au même titre que l'incohérence, l'incomplétude, les données erronées, les données complexes et les présentations imparfaites. Smets a établi une classification de l'imperfection (figure 1) éclatée en trois principaux groupes : imprécision, incohérence et incertitude. Pour l'auteur, l'imprécision est relative au contenu des déclarations. Ces dernières peuvent contenir ou non des erreurs, relever d'un caractère vague, approximatif (le mur mesure plus de deux mètres). L'incohérence regroupe les déclarations conflictuelles, des informations contradictoires (le mur est en pierre alors que le bâtiment est en bois). Quant à l'incertitude, elle est induite soit par des informations objectives (les propriétés de la donnée) soit par des informations subjectives (l'avis de l'observateur).

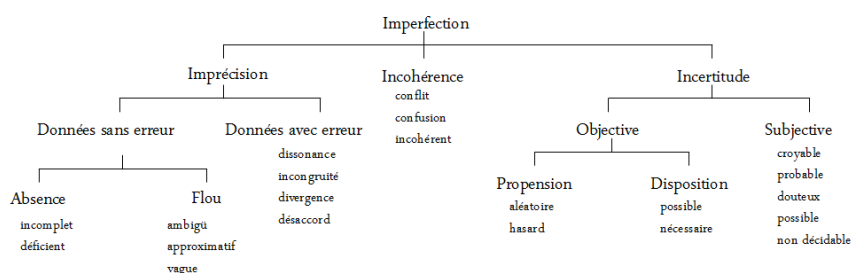


Figure 1. Classification proposée par Smets

Chacune de ces classifications reste discutable et selon le domaine, certaines correspondent mieux à une réalité que d'autres. Néanmoins certaines similitudes peuvent être relevées dans les travaux cités. D'abord, la cohérence (respectivement l'incohérence) se retrouve dans une majorité de classifications. Elle fait référence à l'existence de données contradictoires sur un même objet. Ensuite, l'intervention humaine, représentée dans les classifications par différents termes (interprétation,

prédiction, extrapolation, subjectivité), y est souvent mentionnée comme ayant des conséquences sur la qualité de l'information. Cette intervention peut se situer à différents stades de la collecte des données et induit généralement de l'incertitude, voire de l'imprécision. La difficulté à la quantifier et à la représenter semble réelle. Les outils proposés pour représenter l'imprécision, au sens de Smets, sont majoritairement liés aux théories mathématiques et donnent des résultats intéressants sur les données quantitatives.

Enfin, le dernier élément de comparaison entre les travaux concerne la notion d'incertitude, où les avis sont moins consensuels que sur les deux précédents points. Les classifications qui ont été citées peuvent être classées en deux groupes conformément aux deux définitions énoncées. Un premier groupe fait nettement apparaître l'importance des sources pour la qualité des données à l'image de ce qui se retrouve dans (Thomson *et al.*, 2005) avec la notion d'objectivité, dans (Smets, 1997) avec l'aspect subjectif de l'information ou dans (Skeels *et al.*, 2010) avec la notion de crédibilité. À l'inverse dans le second groupe, les classifications proposées sont davantage orientées vers la représentation de l'incertitude portant sur l'information elle-même.

La plupart du temps, dans le domaine de la géographie ou de l'archéologie, le sens donné à l'incertitude est différent de celui proposé dans (De Runz, 2008) et repris dans cet article. Il correspond davantage au fait d'être dans l'impossibilité de pouvoir attribuer une valeur exacte à une information, plutôt que dans l'existence d'un doute sur la valeur fournie. Pour qualifier cette dualité, le terme souvent employé est imperfection. Pour différencier ces deux formes d'imperfection, (Fisher, 1999) introduit la notion d'objet bien défini et d'objet mal défini. Les objets bien définis sont des objets considérés comme valides par un agent. Ce sont des objets pour lesquels les propriétés connues, même si elles ne sont pas précises, ne peuvent être remises en cause. À l'opposé, les objets mal définis sont ceux pour lesquels une imprécision existe. Dans son approche Fisher associe une théorie de représentation (probabilités, ensembles flous, fonctions de croyance) à chaque type d'imperfection. La théorie des probabilités est généralement retenue pour modéliser la fiabilité d'une source.

Néanmoins si cet outil s'avère efficace pour modéliser l'incertitude des objets bien définis, en revanche, il ne permet pas d'inférer d'autres informations. Il ne donne pas non plus la possibilité de conserver l'origine des différentes valeurs. Selon les domaines d'expertises, les pratiques peuvent différer mais il n'est pas rare que le spécialiste soit obligé de faire un choix pour une des valeurs sans être certain de sa décision. Cette situation génère une donnée qui est alors incertaine. Pour autant, la donnée pourra être précise ou imprécise. Les qualificatifs utilisés dans les différentes typologies pour les données imparfaites ne sont pas à considérer comme exclusifs. Une donnée peut très bien être imprécise et certaine, précise et incertaine ou imprécise et incertaine. L'important est d'être en mesure de l'identifier comme telle et de pouvoir évaluer son effet sur la qualité des données et des décisions qui seront prises. À notre connaissance, il n'y a pas de travaux qui prennent en compte la fiabilité des sources (agents ou ressources) et la conduite de raisonnement sur ce type

d’incertitude. Aussi pour répondre à cet état de fait, cet article se propose d’illustrer comment l’utilisation d’un cadre théorique formel comme les logiques modales avec ses mécanismes de raisonnement peut répondre, dans des contextes bien spécifiques, à cette problématique.

2.2. De la logique classique aux logiques modales

Dans ce paragraphe, nous rappelons les fondamentaux de la logique des prédicats du premier ordre et détaillons les concepts des logiques modales utiles à la compréhension de cet article.

La **logique des prédicats** du premier ordre permet de représenter des connaissances à l’aide de formules basées sur la définition de prédicats n-aires, l’utilisation d’opérateurs logiques (conjonction \wedge , disjonction \vee , négation \neg , implication \rightarrow) et de quantificateurs. Une formule composée uniquement d’un prédicat est une formule atomique ou atome. Les formules non atomiques sont construites à partir d’autres formules, des opérateurs logiques et des quantificateurs universel (\forall) et existentiel (\exists). La restriction aux clauses de Horn permet de disposer de moteurs de raisonnement complet capables de prouver la véracité ou non d’une formule. Une clause de Horn a la forme suivante : $p_1, \dots, p_n \rightarrow p_0$ où chaque p_i est un atome. p_0 représente la tête de la clause et p_1, \dots, p_n en est le corps. Une clause avec une tête et un corps non vides traduit une règle. Cette dernière indique que si les conditions p_1 et ... et p_n sont vraies alors la conclusion p_0 est aussi vraie. Par exemple introduisons les prédicats $m(X, Y)$ et $f(X)$ où X et Y sont des variables représentant des personnes. La formule $m(X, Y)$ exprime que X est une personne et la mère de Y et $f(X)$ spécifie que X est une personne et une femme. La règle $m(X, Y) \rightarrow f(X)$ indique que si X est la mère de Y alors X est une femme. Une clause dont le corps est vide énonce un fait. La formule $f(jade)$ signifie que jade est une personne et une femme. La forme logique d’une règle est $\forall X_1, \dots, X_k (p_0 \vee \neg p_1 \vee \dots \vee \neg p_n)$ où X_i désigne les variables référencées dans les prédicats.

Les **logiques modales** permettent, par l’introduction de nouveaux opérateurs, appelés modaux (\Box et son dual \Diamond), d’enrichir leur pouvoir d’expression et de modéliser des formules plus nuancées. Il existe plusieurs catégories de logiques modales classées selon les interprétations des opérateurs modaux. Par exemple, en logique aléthique les opérateurs sont interprétés par *possible* et *nécessaire*, en logique temporelle on utilise les termes de *parfois* et *toujours*, en épistémique *croit que* et *sait que*, et en déontique les opérateurs interprètent *est permis* et *est obligatoire*. On peut, par exemple, décrire la proposition « Il est possible que jade soit la mère de alain » à l’aide d’une modalité *possible*, ou encore, utiliser dans un contexte épistémique la modalité *sait que* pour affirmer « On sait que jade est une femme ».

La compréhension de ces modalités nécessite une extension de la sémantique de la logique des prédicats du premier ordre basée sur des modèles de Kripke (Gasquet *et al.*, 2013). Ces derniers considèrent que l’univers du discours peut être représenté par différents mondes reliés entre eux par une relation d’accessibilité. Ainsi, en logique

aléthique, une formule φ pourra être interprétée comme *possible* dans le monde actuel, notée $\diamond\varphi$, si elle est vraie dans au moins un des mondes accessibles depuis le monde actuel et fautive ailleurs et comme *nécessaire*, notée $\Box\varphi$, si elle est vraie dans tous les mondes accessibles depuis le monde actuel.

Les **systèmes de déduction** s'appuient sur les axiomes de la logique considérée pour réaliser des déductions à partir des formules décrites et connaître la véracité d'une formule. Ils utilisent les mécanismes de substitution pour remplacer une variable par une autre expression et d'unification (succession de substitutions rendant deux formules égales).

La méthode de résolution SLD (*Selective Linear Definite*) (Lloyd, 1987) est une méthode de preuve complète de la logique des prédicats du premier ordre sur un ensemble P de clauses de Horn. Pour prouver que le but F est vrai, il faut prouver que $P \cup \{\neg F\}$ est inconsistant. Soit $G = \neg F$, la démonstration est faite en déduisant à partir de G une séquence de formules jusqu'à obtenir la clause vide. Les formules G_i sont telles que $G_0 = G$ et pour tout $i \geq 0$, G_{i+1} est obtenue à partir de G_i , de la manière suivante :

- q_m est un atome sélectionné dans le corps de G_i
- $p_1, \dots, p_n \rightarrow p_0$ est une clause du programme P
- θ est l'unificateur le plus général de q_m et p_0
- $G_{i+1} = \rightarrow (q_1, \dots, q_{m-1}, p_1, \dots, p_n, q_{m+1}, \dots, q_r)\theta$

Considérons l'ensemble des propositions du tableau 2 qui sont représentées sous formes clausales à l'aide des constantes stL (pour Saint Leontius), eveque, frejus, des variables X, Y, Z et des prédicats $e(X, Y) : X$ est un Y, $v(X, Z) : X$ a pour ville épiscopale Z, $d(X, Z) : X$ est mort dans la ville Z. Si l'on cherche à déduire que

Tableau 2. Propositions et clauses de Horn

ref	Propositions	formes clausales constituant P
(1)	Saint Leontius est un évêque	$e(\text{stL}, \text{eveque})$
(2)	Saint Leontius a pour ville épiscopale Fréjus	$v(\text{stL}, \text{frejus})$
(3)	Saint Leontius est mort dans sa ville épiscopale	$\neg v(\text{stL}, z) \vee d(\text{stL}, z)$

« stL est mort à Fréjus », il faut montrer que l'ensemble $P \cup \{\neg d(\text{stL}, \text{frejus})\}$ est inconsistant (insatisfiable). La preuve par réfutation est schématisée par un arbre de réfutation (figure 2) dans lequel la clause résultante est reliée aux deux clauses parentes par deux arêtes étiquetées par l'unificateur employé.

Ce principe a été étendu pour raisonner sur des formules de la logique modale. Plusieurs approches ont été proposées, comme la méthode des tableaux sémantiques (Fitting, 1988), la méthode des séquents de Gentzen (Szabo, 1969) ou la programmation logique. La méthode des tableaux sémantiques est comme la SLD résolution, une méthode de preuve par réfutation. La négation de la formule à prouver est décomposée en sous-formules de sorte que si la formule est satisfaite alors les sous-formules qui la composent le sont aussi. La décomposition permet la construction

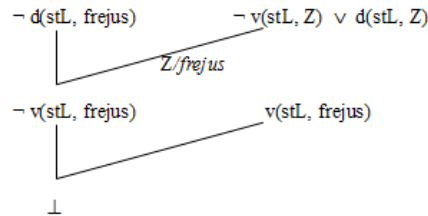


Figure 2. Exemple de résolution SLD : arbre de réfutation

d'un arbre et s'effectue par application de règles d'expansion (de prolongation, de ramification, de double négation, etc.) qui tiennent compte de la sémantique des mondes possibles. Une formule a une preuve si toutes les branches de l'arbre sont fermées, c'est-à-dire qu'une formule et sa négation apparaissent dans chaque branche. En programmation logique, des travaux comme ceux de NGuyen (Nguyen, 2004) et Del Cerro (Del Cerro, 1986) autour du langage de programmation logique Prolog, ont débouché sur la réalisation de moteurs d'inférence capables de prendre en compte les opérateurs modaux. Ils sont discutés dans la section 4.

3. Plateforme pour la gestion de données archéologiques

La variété des acteurs généralement impliqués dans la collecte des données archéologiques nécessite de disposer d'un environnement collaboratif. La plateforme WikiBridge¹ (Chevalier *et al.*, 2010) a été développée pour satisfaire ces objectifs. Elle met, en outre, à disposition des outils pour contrôler la qualité des informations collectées.

3.1. Présentation de l'architecture

Le cœur de WikiBridge est constitué par le couplage entre des bases de données relationnelles et des annotations sémantiques pour produire des connaissances dans un environnement collaboratif évolutif. Ce couplage permet l'extensibilité des modèles de données (schémas) tout en conservant les mécanismes de contrôle de la qualité des informations et le fonctionnement des applicatifs. L'interface utilisateur prend la forme d'un wiki pour décrire des objets, des édifices, des lieux, des personnages au moyen d'une description riche incluant textes, sources bibliographiques, plans, localisations, photographies. Les différents éléments d'une page du wiki peuvent être complétés par des annotations construites à partir d'une ontologie sous la forme de triplets $\langle \text{ sujet, prédicat, objet } \rangle$ (Leclercq *et al.*, 2013).

1. développée puis validée dans le cadre de deux projets : l'ANR CARE *Corpus Architecturae Religiosae Europaeae - IV-X saec.* - ANR-07-CORP-011 (<http://care.tge-adonis.fr>) et le projet de plateforme numérique et mutualiste sur l'usage de la pierre bourguignonne (<http://lapierre.u-bourgogne.fr>).

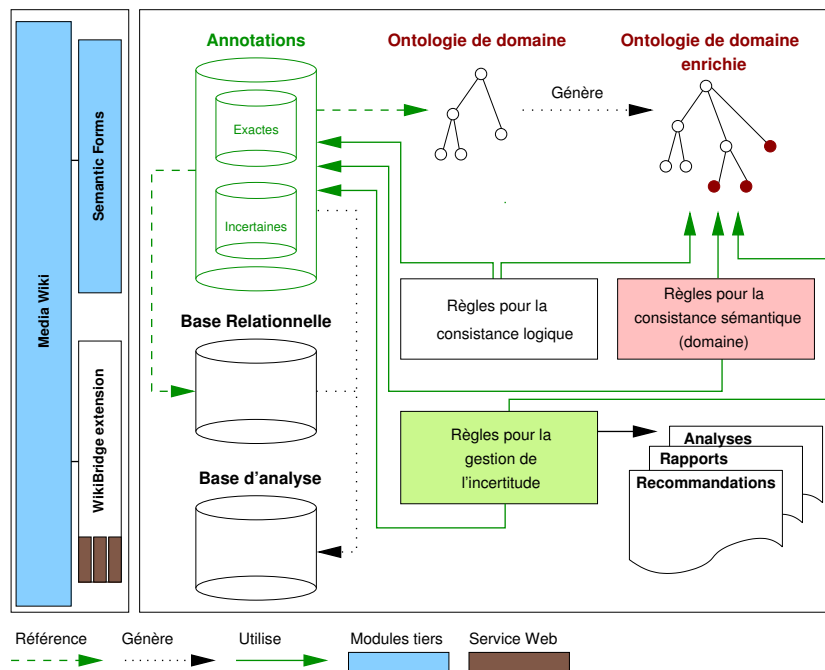


Figure 3. Architecture de WikiBridge

La figure 3 présente les principaux modules de WikiBridge et leurs interactions. Chaque document du corpus se traduit sous la forme d'une page du wiki. Semantic Forms² permet de doter WikiBridge d'une interface d'acquisition de données basée sur des formulaires autorisant l'ajout de données structurées. Cependant les possibilités d'interrogation étant limitées, notamment quand il s'agit de retrouver des informations sur plusieurs pages simultanément, il a été jugé utile de compléter la base de données relationnelle héritée de MediaWiki, d'un modèle d'annotations sémantiques. Ce modèle s'appuie sur une ontologie de domaine. Un assistant aide les utilisateurs à formuler des annotations. Les utilisateurs sélectionnent dans des listes des termes extraits de l'ontologie *via* des requêtes SPARQL prédéfinies et leur associent des propriétés et des valeurs. Un mécanisme de contrôle des annotations incluant la prise en compte du contexte et des contraintes sémantiques a été mis en place. La partie des annotations ne comportant pas d'incertitude (annotations exactes) est ajoutée dans l'ontologie afin de constituer l'ontologie de domaine enrichie et permettre l'utilisation des mécanismes d'inférence pour effectuer les contrôles de consistance. L'autre partie des annotations (annotations incertaines) faisant référence à des connaissances incertaines sont combinées avec les informations

2. https://www.mediawiki.org/wiki/Extension:Semantic_Forms/fr est une extension de MediaWiki pour créer et modifier des pages du wiki *via* des formulaires

extraites de l'ontologie enrichie pour être analysées par les règles de gestion de l'incertitude et rendre compte aux utilisateurs des analyses opérées, des déductions réalisées et des recommandations formulées. Selon les domaines d'expertise, un jeu de règles de gestion de l'incertitude différent peut être fourni par les spécialistes. Le mécanisme de contrôle automatisé et basé sur des règles assure la consistance de ces annotations. Deux types de vérification sont réalisées : l'une logique s'assure que les propriétés associées aux éléments annotés sont cohérentes avec la description ontologique, l'autre sémantique vérifie la cohérence des annotations entre elles (non contradictoires) et la cohérence des annotations par rapport aux règles du domaine (connaissances spécifiques non nécessairement consensuelles). Pour l'analyse spatiale, un ensemble de services Web offre une interaction avec des outils externes et un module de cartographie s'appuyant sur OpenLayers. L'interfaçage avec le système d'information géographique est réalisé grâce à une base de données spécifique (dite base d'analyse) dont le schéma est établi à partir des champs des formulaires et des annotations.

3.2. *Ontologies et modèle d'annotations*

La multiplicité des acteurs impliqués dans le processus de collecte et d'interprétation des données accroît la difficulté d'obtenir un consensus sur ces données et les connaissances représentées. L'usage de bases de connaissances agréées et partagées par un ensemble d'acteurs d'un domaine facilite la modélisation et l'interopérabilité des différents applicatifs qui gèrent ces connaissances. Les ontologies proposent une spécification formelle et explicite d'une conceptualisation partagée d'un domaine (Gruber, 1993). À ce titre, elles sont un moyen facilitant la représentation des connaissances et la contextualisation des données. Les raisonneurs qui leur sont associés prennent en charge la détection d'incohérences logiques et permettent d'inférer de nouvelles connaissances. La modélisation des ontologies est généralement basée sur des langages, sous-ensembles de la logique du premier ordre, issus des logiques de descriptions (Baader *et al.*, 2003) comme OWL³. Dans le domaine du patrimoine culturel, l'ontologie CIDOC-CRM⁴ est une référence qui est devenue un standard. Cette ontologie est largement utilisée dans différents travaux, comme le projet VENUS (Jeansoulin, Papini, 2007) qui propose une représentation d'observations sous-marines à l'aide des concepts et des relations de cette ontologie. Cependant les ontologies n'offrent pas d'opérateurs directement adaptés à la représentation des connaissances imparfaites, seules des plateformes plus complexes mixant plusieurs paradigmes pour la modélisation des connaissances permettent de représenter et d'exploiter cette variété d'information.

Les illustrations proposées dans cette section s'appuient sur un exemple concret. La figure 4 présente deux ressources provenant de Wikipédia ayant comme sujet

3. Ontology Web Language (<http://www.w3.org/TR/owl-features/>)

4. Comité International pour la Documentation - Conceptual Reference Model qui a été élaborée depuis 1994 et publiée en 2006 par l'ISO (<http://www.cidoc-crm.org>)

identifiée par une URL et les annotations utilisent cette URL comme préfixe pour identifier les éléments annotés. Par la suite et pour simplifier l'écriture, les annotations relatives au document (b) de la figure 4 seront identifiées de manière courte par #stL (abréviation pour Saint Leontius ou Léonce de Fréjus).

Une annotation est un triplet $\langle s, p, o \rangle$ liant un sujet s , un prédicat p et un objet o . Le sujet correspond à l'élément que l'on souhaite annoter. Le couple (p, o) représente le contenu de l'information que l'on souhaite ajouter au sujet. Le prédicat exprime la relation entre le sujet et l'objet. Il donne du sens à l'annotation et peut être un concept ou une propriété de l'ontologie. À titre d'exemple, l'annotation générée automatiquement à partir des éléments fournis dans l'interface pour représenter le fait que "Saint Léonce est évêque" est la suivante : $\langle \#stL, estUn, Evêque \rangle$.

Le corpus permet généralement d'extraire de nombreuses informations relatives à un même sujet. Certaines sont directement en relation avec le sujet comme par exemple "Saint Léonce est un évêque" ou "Saint Léonce est fêté le 1^{er} décembre". Elles qualifient le sujet identifié "Saint Léonce" (#stL) alors que d'autres peuvent qualifier l'objet en liaison avec le sujet. C'est le cas du soupçon émis par Célestin 1^{er} envers Saint Léonce (extrait fiche 4b). Dans ce cas Saint Léonce est soupçonné d'un méfait. Ce dernier est sa croyance en liaison avec le semi-pélagianisme. Il a été dénoncé par Célestin 1^{er}. Les deux affirmations précédentes qualifient le méfait dont Saint Léonce est soupçonné. Dans ce cas il est primordial de ne pas perdre le lien avec le sujet (Saint Léonce).

Pour modéliser tous ces cas de figures, trois types de représentation sous forme d'annotations ont été proposés :

1. une annotation *simple* permet d'annoter un sujet s représenté par son identifiant en lui associant un prédicat p et un objet o . Par exemple, $\langle \#stL, estUn, Evêque \rangle$ est une annotation simple ;

2. une annotation *réflexive* est représentée par l'imbrication de plusieurs triplets. Elle permet de mettre en relation plusieurs annotations liées au même sujet mais qualifiant l'objet du sujet. Dans l'exemple, ce cas s'applique au soupçon émis par Célestin 1^{er} à propos de Saint Léonce. L'annotation réflexive correspondante est la suivante :

$$\langle \#stL, estSoupçonnéDe, acteStL \\ \langle \langle acteStL, aPourNature, croyanceSemiPelagianisme \rangle, \\ \langle acteStL, aPourDate, 431 \rangle, \\ \langle acteStL, estDénoncéPar, Célestin1er \rangle \\ \rangle \\ \rangle$$

Les triplets ayant comme sujet *acteStL* ne doivent pas être dissociés du sujet initial. Ils constituent un complément d'information sur l'objet *acteStL* relié au sujet #stL par le prédicat *estSoupçonnéDe*;

3. une annotation *complexe* est un ensemble d'annotations portant sur le même sujet et qui peuvent être simples ou réflexives. Pour l'exemple de Saint Léonce, une

annotation complexe pourrait être composée de l'ensemble suivant d'annotations :

```

<
  < #stL, estUn, Evêque >,
  < #stL, estSoupçonnéDe, acteStL
    < acteStL, aPourNature, croyanceSemiPelagianisme >,
    < acteStL, aPourDate, 431 >,
    < acteStL, estDénoncéPar, Célestin 1er >
  >
  >,
  < #stL, aFête, '01-12' >
>

```

Dans ce modèle d'annotation, le prédicat, qui est un terme de l'ontologie, ne peut pas être le sujet d'une annotation car la connaissance consensuelle sur cet élément est exprimée dans l'ontologie.

La suite présente les pistes de travail explorées pour compléter les outils de raisonnement présents dans la plateforme WikiBridge par des outils de raisonnement sur des données et des connaissances incertaines. Ces dernières sont représentées par des annotations accompagnées d'opérateurs modaux et exploitées avec la logique des prédicats.

4. Incertitude : de la théorie à l'implémentation

Le terme incertitude utilisé dans cet article est lié à la présence de doute dans la description de la connaissance. Les caractéristiques, tant spatiales que temporelles, des objets sont définissables par des sous-ensembles de valeurs ou des intervalles. L'incertitude sur la validité d'une position spatiale ou temporelle peut avoir plusieurs origines parmi lesquelles deux nous intéressent plus particulièrement : la source de l'information et l'interprétation humaine. Une source (document, mesure, enregistrement, etc.) peut être considérée comme fiable à un instant t_1 et se révéler postérieurement non fiable à un instant t_2 . Parallèlement à la source, l'expertise peut conduire à des résultats qui peuvent contenir une part d'incertitude. Le spécialiste, au regard de son expérience, peut inférer de nouvelles connaissances sans pour autant être certain de ce qu'il avance. Pouvoir représenter ce type d'information est un des objectifs du travail présenté.

4.1. Raisonnement dans les systèmes de logique modale

De nombreux travaux ont proposé des extensions de la programmation logique avec des approches différentes mais seuls le groupe de Fariñas del Cerro avec le système Molog (Del Cerro, 1986) et Linh Anh NGuyen avec MProlog (Nguyen, 2003) ont abouti à des implémentations. Si les deux systèmes optent pour la même démarche, celui de Nguyen n'impose aucune restriction quant à l'utilisation des opérateurs modaux. Il est basé sur le modèle de Kripke et intègre le calcul de résolution SLD.

Ce modèle étendu prend en compte l'axiome de base K^5 ainsi que d'autres axiomes qui seront détaillés dans les exemples.

Une formule modale est de la forme $[\Delta] \varphi$ où Δ est une liste d'opérateurs modaux et φ une clause de Horn pouvant comporter des opérateurs modaux.

Les formules proposées dans le tableau 3 illustrent différentes possibilités d'expression des modalités et montrent qu'un opérateur modal peut indifféremment porter sur une clause ou sur un prédicat, dans le corps ou la tête de la clause.

Tableau 3. Exemples de formules modales

ref	formule modale	interprétation
φ_1	$\diamond m(\text{jade}, \text{alain})$	possible que m(jade, alain)
φ_2	$\Box(m(X, Y) \rightarrow \Box f(X))$	nécessaire si m(X, Y) alors nécessaire f(X)
φ_3	$\Box(m(X, Y), \Box f(X) \rightarrow p(X, Y))$	nécessaire que si m(X, Y) et nécessaire f(X) alors p(X, Y)
φ_4	$\diamond p(X, Y) \rightarrow \Box e(Y)$	si possible p(X, Y) alors nécessaire e(Y)

Dans (Nguyen, 2003) l'auteur détaille l'extension du calcul de la résolution SLD aux logiques modales dans une approche fidèle à la programmation logique classique. Lors de la construction de l'univers des mondes, pour rendre compte qu'une formule modale $\diamond\varphi$ est présente dans le monde actuel w_0 il convient de connecter, au monde actuel, un monde w_1 et d'ajouter φ dans w_1 . Le fait que φ soit vrai dans le monde w_1 est traduit sous la forme suivante $\langle\varphi\rangle\varphi$. De la même manière, tout ce qui pourra être déduit dans le monde w_1 sera étiqueté de $\langle\varphi\rangle$. En effet, cette étiquette indique que les propositions déduites le sont dans un monde w_1 où la proposition φ est vraie. Cela reflète la notion de possible : vrai dans w_1 mais pas forcément vrai ailleurs dans un autre monde accessible depuis w_0 . À l'inverse, la prise en compte d'une formule $\Box\varphi$ dans le monde actuel w_0 va se traduire par l'ajout de φ dans tous les mondes accessibles depuis w_0 .

Pour illustrer les étapes du raisonnement, nous proposons de prendre l'exemple de la résolution du but $G = \leftarrow \Box e(Y)$ dans la logique monomodale KD^6 . Le processus de résolution est schématisé dans la figure 6. Il s'appuie sur le programme P constitué des formules incluant les modalités présentées dans le tableau 3.

En considérant G et φ_4 nous dérivons un nouveau but $G_1 = \leftarrow \diamond p(X, Y)$. À partir de G_1 et φ_3 , nous déduisons le but $\diamond(m(X, Y) \wedge \Box f(X))$ qui est constitué d'une séquence de deux opérateurs modaux $\diamond \Box f(X)$. Or dans les logiques monomodales, la séquence d'opérateurs modaux a une longueur au plus égale à un. Pour permettre néanmoins la résolution de la formule $\diamond p(X, Y)$, cette dernière va être étiquetée par $\langle Z \rangle$ où Z est une variable libre (non quantifiée) pour devenir $\langle Z \rangle p(X, Y)$. L'application de φ_3 et K à ce but étiqueté va donner le nouveau but $G_2 = \leftarrow \langle Z \rangle m(X, Y) \wedge \langle Z \rangle \Box f(X)$. Le

5. Axiome K : $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$ où φ et ψ sont des formules

6. Axiomes K et D

axiome D : $\Box\varphi \rightarrow \diamond\varphi$ se traduit par si φ est nécessaire alors φ est possible

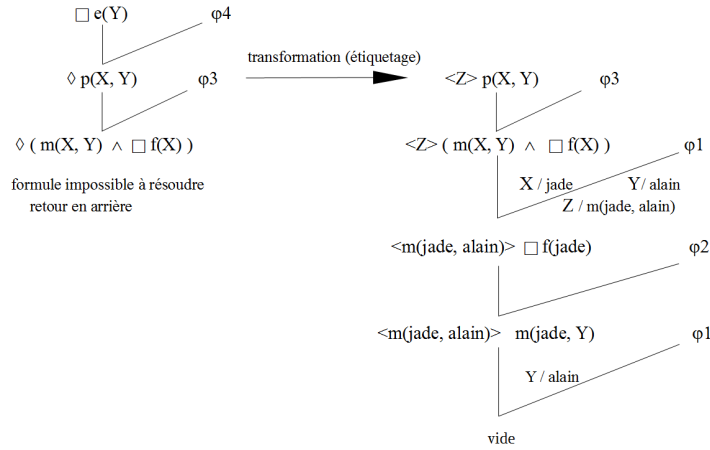


Figure 6. Exemple de résolution étendue à la logique monomodale KD

rapprochement de G_2 et φ_1 avec l’hypothèse $\langle m(jade, alain) \rangle m(jade, alain)$ fournit $G_3 = \leftarrow \langle m(jade, alain) \rangle \Box f(jade)$ grâce à l’unificateur $X/jade, Y/alain, Z/m(jade, alain)$. De G_3 et φ_2 nous obtenons $G_4 = \leftarrow \langle m(jade, alain) \rangle m(jade, Y)$. En appliquant φ_1 à G_4 nous dérivons $G_5 = \leftarrow \langle m(jade, alain) \rangle m(jade, alain)$. Et finalement, en unifiant G_5 et φ_1 nous arrivons à un ensemble de formules vide. La solution au but recherché $\leftarrow \Box e(Y)$ est donc alain puisque les formules inférées sont toutes vraies dans l’hypothèse que $m(jade, alain)$. Cet exemple développe les principaux mécanismes mis en œuvre par l’interpréteur développé par NGuyen pour résoudre le but, lesquels seront très utiles pour comprendre les cas d’utilisation traités.

4.2. Modélisation et implémentation de différentes formes d’incertitude

Notre contribution porte sur la modélisation de certaines formes d’incertitude avec l’utilisation d’une théorie basée sur la logique modale et la logique des prédicats. Cette section décrit, à travers deux exemples, les nouvelles connaissances inférées lors du raisonnement par l’existence de données incertaines.

Pour illustrer les propos, la ressource (extrait fiche 4a) comportant le plus d’imprécisions servira de support. Dans cette fiche, les informations pour lesquelles aucune incertitude n’existe sont : (i) « St Leontius est évêque de Fréjus en Provence », (ii) « il est mort dans sa ville épiscopale », (iii) « sa fête est le 1^{er} décembre ». L’affirmation « il est né vers la fin du 4^e siècle » est en contradiction avec l’année de naissance fournie (419). Elle pourrait être traitée comme le seront les trois dates mais pour les besoins de l’illustration, l’année de naissance sera préférée (iv). Pour ce qui est des informations apparaissant comme incertaines, seul le lieu de naissance est concerné : (v) « il est probablement né à Nîmes ». Enfin, la présence de trois dates différentes (vi) pour la mort de Saint Leontius soulève des interrogations. Pour représenter cet état de fait, plusieurs solutions sont envisageables : ne modéliser

aucune donnée au risque de les perdre, modéliser une seule d’entre elles en effectuant un choix, ou mémoriser toutes ces données. C’est cette dernière alternative qui est retenue. Le raisonnement étant le même pour deux ou trois dates, il sera développé avec les deux premières dates seulement. L’extrait ne précise pas que St Leontius est mort à Fréjus. Néanmoins cette information sera automatiquement déduite. En logique des prédicats du premier ordre il est syntaxiquement possible d’énoncer trois faits avec trois valeurs mais dans la réalité disposer, pour un événement, de trois dates relève d’une incohérence. Il y a forcément des valeurs non correctes parmi celles fournies et tout au plus une valeur correcte. Disposer des opérateurs modaux offre la possibilité de représenter ces valeurs comme des données incertaines et non plus comme incohérentes.

Pour être en mesure de raisonner en logique des prédicats sur les triplets représentant les annotations, ces derniers sont transcrits sous forme clausale. Un prédicat *an* (abrégé de annotation) d’arité trois ayant pour valeurs les éléments du triplet vient se substituer à l’annotation. Par exemple, le triplet $\langle \#stL, estUn, Evêque \rangle$ s’écrit sous forme clausale $an(\#stL, estUn, evêque)$.

Les deux cas suivants proposent différentes exploitations des mêmes données à travers deux modalités : aléthique et épistémique.

Cas 1 : implémentation en logique modale aléthique.

En logique modale aléthique, les modalités qualifient la véracité des données sans que la fiabilité de la personne à l’origine de l’information soit prise en compte. Les annotations peuvent être fournies par la même personne à partir de plusieurs ressources ou par plusieurs personnes à partir d’une ou de plusieurs ressources sans qu’il soit utile de différencier la provenance. L’identité de la personne qui annote les documents n’est pas conservée. Le tableau 4 reprend les informations de la fiche a (figure 4) et propose leur formulation en utilisant la logique modale aléthique. La portion de texte référencée précédemment par (i) donne lieu à trois formules numérotées $f1$, $f2$ et $f3$. Pour montrer la prise en compte des incertitudes (nécessaire

Tableau 4. Formules en logique modale aléthique

Réf au texte	Formule avec ou sans modalité	Identification
(i)	$an(\#stL, estUn, evêque)$	f1
	$an(\#stL, aPourVilleEpiscopale, 'Fréjus')$	f2
	$an('Fréjus', estLocaliseEn, 'Provence')$	f3
(ii)	$an(\#stL, aPourVilleEpiscopale, V^a) \rightarrow an(\#stL, estMortA, V)$	r1
(iii)	$an(\#stL, aFete, '01-12')$	f4
(iv)	$\square an(\#stL, estNeEn, 419)$	$\square f5$
(v)	$\diamond an(\#stL, estNeA, 'Nimes')$	$\diamond f6$
(vi)	$\diamond an(\#stL, estMortEn, 488)$	$\diamond f7$
	$\diamond an(\#stL, estMortEn, 443)$	$\diamond f8$

a. Une valeur identifiée par un nom commençant par une majuscule représente une variable. La variable V représente la ville où #stL est mort et qui est aussi sa ville épiscopale.

et son dual possible) dans le raisonnement, nous rajoutons aux formules précédentes du tableau 4 les deux règles suivantes :

- r2 : $\text{an}(\text{Suj}, \text{Pred}, \text{Obj}) \rightarrow \Box \text{an}(\text{Suj}, \text{Pred}, \text{Obj})$ ⁷
- \Box r3 : $\Box (\text{an}(\text{Suj}, \text{estMortEn}, D_m), \text{an}(\text{Suj}, \text{estNeEn}, D_n), \text{Age} = D_m - D_n \rightarrow \text{an}(\text{Suj}, \text{estMortAge}, \text{Age}))$

La règle r2 permet de traiter les formules sans modalité comme des formules nécessairement vraies. La règle se lit « Si la formule est vraie alors il est nécessaire qu'elle soit vraie ». En d'autres termes, ce qui est vrai dans un monde doit l'être aussi dans tous les mondes accessibles depuis ce monde. Un seul point de vue est considéré. Le fait f2 et la règle r1 (4^{ème} ligne du tableau 4) permettent de déduire que Saint Leontius est mort à Fréjus. La règle r2 rend ce nouveau fait vrai non seulement dans le monde actuel mais dans tous les mondes accessibles depuis le monde actuel. Elle permet de propager les formules à tous les mondes.

La règle r3 permet de déduire l'âge d'une personne au moment de son décès en fonction de son année de mort (D_m) et de son année de naissance (D_n). La modalité *nécessaire* (\Box) s'applique à l'ensemble de la formule r3. Cette formule sera donc vraie dans tous les mondes accessibles depuis le monde actuel.

La figure 7 schématise l'univers des mondes créé à partir des annotations au sujet de Saint Leontius. Le monde actuel est w_0 . Il renferme toutes les formules du tableau 4 à l'exception de $\Diamond f_6$ qui ne présente pas d'intérêt pour notre exemple. À partir de w_0 , deux mondes sont accessibles, w_1 et w_2 , puisque deux faits sont décrits comme possibles (f_7 , f_8). L'unification de la formule r3 dans le monde w_1 où la formule f_7 , stipulant que Saint Leontius est mort en 488, est considérée comme vraie va inférer le fait $\text{an}(\#stL, \text{estMortAge}, 69)$ (Saint Leontius est mort à l'âge de 69 ans). Ce raisonnement est reproduit à l'identique dans le monde w_2 pour déduire $\text{an}(\#stL, \text{estMortAge}, 24)$.

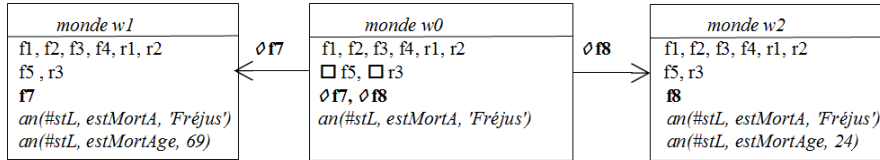


Figure 7. Extrait de la représentation des mondes

À des fins de vérification de cohérence, il est intéressant d'obtenir toutes les données incertaines et notamment celles qui ont été déduites. La résolution du but $\Diamond \text{an}(\text{Suj}, \text{Pred}, \text{Obj})$ va répondre à cette attente. La requête initiale $\Diamond \text{an}(\text{Suj}, \text{Pred}, \text{Obj})$ dans le monde actuel va restituer les formules qui étaient déjà incertaines (f_7 , f_8) auxquelles vont être rajoutées les deux nouvelles propositions déduites et notées

7. Suj représente le sujet, Pred le prédicat et Obj l'objet. Le prédicat relie le sujet à l'objet. Il représente un concept ou une propriété de l'ontologie.

en italique dans la figure 7. Chacune sera étiquetée du fait dont elle dépend. Quant au fait relatif au lieu de mort, il est déduit de manière identique dans tous les mondes, puisqu'il est basé sur des connaissances vraies dans tous les mondes.

L'exemple ci-dessus proposé à partir des dates montre comment la déduction logique à partir de données incertaines peut inférer de nouvelles connaissances tout aussi incertaines. La visualisation des déductions opérées va permettre à l'expert du domaine de s'interroger sur leur véracité, la nécessité ou non de les conserver en l'état ou de les transformer. Il pourra ainsi faire autant de simulations qu'il le souhaite pour modéliser au plus juste le monde réel. Cependant certaines fois ce choix restera délicat et d'autres besoins peuvent se faire sentir notamment la possibilité de différencier les sources (ressources, agents).

Cas 2 : implémentation en logique multimodale épistémique. La multiplicité des acteurs impliqués dans la collecte des données peut à un moment ou à un autre justifier de pouvoir différencier les sources. Dans ces situations particulières, quand les circonstances justifient un intérêt à conserver toutes les informations pour opérer des déductions différenciées, il est nécessaire de passer à une autre forme de logique : la logique multimodale $K_{(m)}$ avec m paires d'opérateurs modaux \Box_i et \Diamond_i , $i \in [1..m]$. Un extrait des axiomes relatifs à cette logique est représenté dans le tableau 5. Dans ce dernier, la formule modale $\Box_i\varphi$ se lit « l'agent i sait que φ est vraie » et $\Diamond_i\varphi$ se lit « l'agent i croit que φ est vraie » où φ est elle-même une formule.

Tableau 5. Extrait des axiomes de la logique multimodale $K_{(m)}$

Nom	Règle	Commentaire
D	$\Box_i\varphi \rightarrow \neg\Box_i\neg\varphi$	consistance
4s	$\Box_i\varphi \rightarrow \Box_j\Box_i\varphi$	introspection positive
5s	$\neg\Box_i\varphi \rightarrow \Box_j\neg\Box_i\varphi$	introspection négative

La plateforme collaborative a pour objectif de constituer une base de connaissances communes au sein d'un groupe de confiance où chacun peut exprimer ses connaissances (certaines ou incertaines) et raisonner en exploitant aussi celles des autres. Ayant besoin de représenter une connaissance individuelle, nous avons retenu la logique $KD4_s5_s$. Elle comprend l'axiome K ainsi que tous ceux détaillés dans le tableau 5.

Pour illustrer les possibilités offertes par cette logique, nous proposons de considérer trois sources différentes nommées S1, S2 et S3 portant sur le même sujet (Saint Leontius). Le cas envisagé est la saisie des annotations relatives à la source S3 sachant que les annotations des sources S1 et S2 sont déjà présentes dans la base de connaissances (tableau 6). Les valeurs des modalités pour S1 et S2 ont été choisies de manière à couvrir un maximum de combinaisons possibles. La source S3 correspond exactement au contenu de l'extrait décrit en figure 4a. Pour faciliter la compréhension du raisonnement en logique multimodale épistémique, seules les formules indispensables issues du tableau 4 sont reprises dans le tableau récapitulatif 6.

Tableau 6. Formulation des données en $KD4_s5_s$

Identification	Formule	S1	S2	S3
f2	$\text{an}(\#stL, aPourVilleEpiscopale, 'Fréjus')$	\square_1	\diamond_2	\square_3
f5	$\text{an}(\#stL, estNeEn, 419)$	\square_1	\square_2	\square_3
f6	$\text{an}(\#stL, estNeA, 'Nimes')$	\square_1	\square_2	\diamond_3
f7	$\text{an}(\#stL, estMortEn, 488)$	\square_1		\diamond_3
f8	$\text{an}(\#stL, estMortEn, 443)$		\square_2	\diamond_3
r1	$\text{an}(\#stL, aPourVilleEpiscopale, V)$ $\rightarrow \text{an}(\#stL, estMortA, V)$	\diamond_1	\diamond_2	\diamond_3
r3	$\text{an}(S, estMortEn, D_m), \text{an}(S, estNeEn, D_n),$ $\text{Age} = D_m - D_n \rightarrow \text{an}(S, estMortAge, \text{Age})$	\square_1	\square_2	\square_3

Chaque ligne du tableau retranscrit les croyances des sources sur une même formule. Elle comprend l'identification et le contenu de la formule ainsi que dans les trois dernières colonnes, les croyances attribuées aux trois différentes sources (S1, S2 et S3). Les symboles \square_1 (respectivement \diamond_1) se lisent « la source S1 sait que la formule correspondante est vraie » (respectivement « la source S1 croit que la formule correspondante est vraie »). La première ligne du tableau fait état que S1 et S3 savent que la ville épiscopale de Saint Leontius est Fréjus (f2), alors que S2 est incertain. À l'inverse, toutes les sources savent que Leontius est né en 419 (f5) et que le calcul de l'âge d'un individu au moment de son décès est la différence entre l'année de mort et celle de naissance (r3). Pour le fait (f6), S1 et S2 savent qu'il est vrai alors que la source S3 le dit incertain. Concernant la date de sa mort, la base de connaissances renferme déjà deux dates données par S1 et S2, qui se retrouvent aussi dans S3. Pour conserver cette information plutôt que de la passer sous silence, chacune des formules va être qualifiée de possible pour S3. Enfin pour la règle relative au lieu de sa mort (r1), les trois sources la croient vraie. Ainsi représentées, les données peuvent être consultées en tenant compte de leur fiabilité.

Dans l'optique d'un travail collaboratif, il semble intéressant d'utiliser cette logique multimodale pour relier la donnée à sa source et disposer de moyens pour vérifier des convergences ou des divergences. Avec les modalités offertes par la logique $KD4_s5_s$, un raisonneur comme celui développé par NGuyen peut extraire à la demande les annotations souhaitées. En dehors des interrogations classiques (celles d'un même sujet, relatives à une source spécifique, etc.) la demande peut concerner les annotations confirmées, probables ou possibles. Les confirmées sont celles considérées comme vraies par toutes les sources. Seule l'annotation (f5) répond à ce critère. Les annotations probables correspondent à celles vraies pour toutes les sources à l'exception d'une spécifiée. Les annotations possibles sont celles pour lesquelles au moins une source les considère comme incertaines. Avec ce type de modalités, il devient possible de déduire des informations, opérer des recoupements afin d'aider l'archéologue à juger de la qualité des données collectées.

5. Conclusion

La modélisation et le raisonnement sur des données incertaines sont des tâches complexes qui nécessitent l'usage d'environnement multi-paradigmes. Le travail présenté dans l'article s'intègre dans une plateforme collaborative de gestion de données archéologiques qui comporte un modèle d'annotation sémantique à base ontologique acceptant l'extension du modèle de données. Nos travaux sur le raisonnement en présence d'incertitude rejoignent et complètent les travaux sur la qualité de données déjà développés au sein de la plateforme.

L'approche décrite dans cet article présente et illustre la logique monomodale aléthique avec les opérateurs modaux de *possibilité* et de *nécessité*, et la logique multimodale épistémique qui élargit la précédente avec la notion d'agents qui peuvent avoir des croyances différentes. Les logiques modales sont utilisées pour représenter les connaissances incertaines mais aussi pour définir des règles de gestion de ces données permettant de faire émerger des connaissances utiles aux archéologues.

Les travaux et l'expérimentation réalisés ont permis d'envisager plusieurs perspectives de développement en relation avec différents points de la proposition. La première est liée aux mécanismes de recommandation qui peuvent être ajoutés à la plateforme afin de contrôler la qualité des données, y compris multi-sources, mais aussi pour confronter les points de vue ou les interprétations des archéologues, historiens, etc. La seconde, plus théorique, concerne la spécification d'un langage d'interrogation de données avec des opérateurs de plus haut niveau permettant de prendre en compte l'incertitude et d'interroger les annotations exactes et incertaines.

Remerciements

Ces travaux ont été partiellement financés par le CNRS au travers du projet PEPS LogModArchéo en 2013 et 2014

Bibliographie

- Baader F., Calvanese D., McGuinness D. L., Nardi D., Patel-Schneider P. F. (Eds.). (2003). *The description logic handbook: Theory, implementation, and applications*. New York, NY, USA, Cambridge University Press.
- Balbani P., Del Cerro L. F., Tinchev T., Vakarelov D. (1997). Modal logics for incidence geometries. *Journal of Logic and Computation*, vol. 7, n° 1, p. 59–78.
- Bennett B. (1996). Modal logics for qualitative spatial reasoning. *Logic Journal of IGPL*, vol. 4, n° 1, p. 23–45.
- Benthem J. van, Bezhanishvili G. (2007). Modal logics of space. In *Handbook of spatial logics*, p. 217–298. Springer.
- Chevalier P., Leclercq E., Millereux A., Sapin C., Savonnet M. (2010). WikiBridge: a Semantic Wiki for Archaeological Applications. In *Proceedings of the 38th conference on computer applications and quantitative methods in archaeology (CAA)*, p. 193–196.

- Del Cerro L. F. (1986). Molog: A system that extends prolog with modal logic. *New Generation Computing*, vol. 4, n° 1, p. 35–50.
- De Runz C. (2008). *Imperfection, temps et espace: modélisation, analyse et visualisation dans un sig archéologique*. Thèse de doctorat. Université de Reims Champagne-Ardenne.
- Doukari O., Jeansoulin R. (2007). Space-contained conflict revision, for geographic information. In *The european information society*, p. 335–347. Springer.
- Fisher P. F. (1999). Models of uncertainty in spatial data. *Geographical information systems*, vol. 1, p. 191–205.
- Fitting M. (1988). First-order modal tableaux. *Journal of Automated Reasoning*, vol. 4, n° 2, p. 191–213.
- Garion C., Roussel S., Cholvy L. (2010). Une logique modale pour raisonner sur la cohérence et la complétude de réglementations. *Revue d'intelligence artificielle*, vol. 24, n° 3, p. 267–290.
- Gasquet O., Herzig A., Said B., Schwarzenrüber F. (2013). *Kripke's worlds: An introduction to modal logics via tableaux*. Springer Science & Business Media.
- Gershon N. (1998). Visualization of an imperfect world. *Computer Graphics and Applications, IEEE*, vol. 18, n° 4, p. 43–45.
- Gruber T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, vol. 5, n° 2, p. 199–220.
- Héon M., Nkambou R. (2013). G-OWL: Vers un langage de modélisation graphique, polymorphe et typé pour la construction d'une ontologie dans la notation OWL. In *24èmes journées francophones d'ingénierie des connaissances*.
- Hué J., Serayet M., Drap P., Papini O., Würbel E. (2011). Underwater archaeological 3D surveys validation within the removed sets framework. In *Symbolic and quantitative approaches to reasoning with uncertainty*, p. 663–674. Springer.
- Jeansoulin R., Papini O. (2007). Underwater archaeological knowledge analysis and representation in the venus project. In *21th CIPA international symposium, anticipating the future of the cultural past*, p. 6.
- Leclercq E., Savonnet M. (2012). Système d'information pour la production de connaissances. l'approche wiki sémantique. *Ingénierie des Systèmes d'Information*, vol. 17, n° 3, p. 143–166.
- Leclercq E., Savonnet M., Troya-Galvis A., Buttner S. (2013). Investigating a multi-paradigm system for the management of archaeological data: Corpus lapidum burgundiae. In *Digital heritage international congress*, vol. 1, p. 679–682.
- Lloyd J. W. (1987). *Foundations of logic programming. symbolic computation: Artificial intelligence*. Springer-Verlag.
- Nguyen L. A. (2003). A fixpoint semantics and an sld-resolution calculus for modal logic programs. *Fundamenta Informaticae*, vol. 55, n° 1, p. 63–100.
- Nguyen L. A. (2004). The modal logic programming system mprolog. In *Logics in artificial intelligence*, p. 266–278. Springer.

- Piolle G., Demazeau Y. (2008). Une logique pour raisonner sur la protection des données personnelles. In *16e congrès francophone sur la reconnaissance de formes et l'intelligence artificielle (RFIA '08)*, p. 8.
- Skeels M., Lee B., Smith G., Robertson G. G. (2010). Revealing uncertainty for information visualization. *Information Visualization*, vol. 9, n° 1, p. 70–81.
- Smets P. (1997). Imperfect information: Imprecision and uncertainty. In *Uncertainty management in information systems*, p. 225–254. Springer.
- Szabo M. E. (1969). *The collected papers of gerhard gentzen* (vol. 160). North-Holland Amsterdam.
- Thomson J., Hetzler E., MacEachren A., Gahegan M., Pavel M. (2005). A typology for visualizing uncertainty. In *Electronic imaging 2005*, p. 146–157.