UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE : STITS

Laboratoire *Images, Signaux et Systèmes Intelligents (LiSSi)*

DISCIPLINE INFORMATIQUE

THÈSE DE DOCTORAT

soutenue le 20/11/2014

par

Olfa MABROUKI

Approche sémantique de gestion des politiques de la vie privée. Application au contrôle des interactions entre les usagers et les environnements d'intelligence ambiante.

Directeur de thèse :	Yacine AMIRAT	Professeur (Universite Paris-Est Creteil)
Composition du jury :		
Président du jury :	Nicole LEVY	Professeur (CNAM)
Rapporteurs :	Patrick REIGNIER	Professeur (Université Joseph Fourier)
	Saïd TAZI	Maître de Conferences, HDR (Université Toulouse 1)
Examinateurs :	Abdelghani CHIBANI	Maître de Conference (Université Paris-Est Créteil)
	Yacine BELLIK	Maître de Conference, HDR (Universite Paris-Sud)

L'utilisation massive des services de l'intelligence ambiante dans la vie quotidienne pose des défis technologiques et sociétaux majeurs qui concernent la protection de la vie privée des usagers. Pour y faire face, des politiques doivent être établies et des contraintes d'ordres éthique et juridique doivent être prises en compte pour contrôler efficacement l'utilisation des services ambiants. Ces derniers opèrent en général sur des objets relatifs à la sphère privée de l'utilisateur, tels que des objets physiques de type robots compagnons, des caméras, des capteurs de localisation ou des objets logiques tels que l'agenda, le carnet de contacts ou le dossier médical personnel.

La problématique de protection de la vie privée doit être prise en compte obligatoirement dès la phase de conception des services ambiants. Il s'agit d'appliquer le principe de « Privacy by Design » qui vise à offrir aux concepteurs une méthodologie et des modèles d'architectures adéquats pour rendre leurs services sensibles à la vie privée.

Plusieurs verrous doivent être considérés dans ce cadre. Le premier, de type architectural, concerne l'interopérabilité sémantique des systèmes de protection de la vie privée. En effet, une politique et un système centralisé pour la gestion de la sécurité et de la vie privée, sont généralement employés dans les systèmes d'information des entreprises. Comparé à ces systèmes, un système intelligent ambiant est caractérisé par des interactions de type "pair à pair" entre des terminaux clients et des services hétérogènes. Chaque pair dispose de sa propre politique de gestion de la vie privée, et de son propre mécanisme de contrôle.

Dans ce contexte, il est difficile d'envisager un organe central dans l'environnement ambiant permettant d'assurer l'interopérabilité entre les politiques des différents pairs. Le deuxième défi concerne l'adaptation des politiques aux préférences et aux contextes des usagers. Ces derniers sont en perpétuel mouvement, changent de contexte souvent et interagissent à distance avec d'autres usagers et utilisent des services fournis par des systèmes hétérogènes à travers le web. Par conséquent, ces usagers doivent disposer d'outils simples leur permettant de définir des politiques adaptatives à leurs contextes.

La protection de la vie privée couvre plusieurs aspects correspondant aux modalités normatives suivantes: les interdictions, les autorisations et les obligations. Ces dernières sont associées à des actions de contrôle concernant, d'une part, la divulgation, l'anonymisation ou l'obfuscation de données personnelles pendant les opérations de transmission ou de stockage de ces données, et d'autre part, l'ouverture, la restriction ou la désactivation des d'accès aux services capteurs ou actionneurs présents dans l'environnement ambiant de l'usager. Par exemple, désactiver l'accès à la caméra du robot compagnon quand ce dernier accompagne l'usager dans la salle de bain. La couverture fonctionnelle complète de tous ces aspects dans la mise en oeuvre de chaque service d'intelligence ambiante est un enjeu important qui nécessite beaucoup de temps et d'attention lors de la conception du service.

Les approches de protection de la vie privée proposées dans l'état de l'art, s'accordent sur le besoin de disposer d'architectures de contrôle d'accès standards et de langages de politiques de gestion de la vie privée qui peuvent être interprétés de la même manière par des systèmes hétérogènes. Ces approches peuvent être classées en trois catégories principales. La première concerne des langages de marquage permettant de définir des règles de consentement pour l'accès ou l'exploitation de données échangées avec des serveurs web. La deuxième catégorie concerne l'utilisation du langage de politiques XACML pour exprimer des politiques de gestion de la vie privée basées sur des modèles de contrôle d'accès classiques RBAC et ABAC. Pour pallier aux problèmes d'interopérabilité sémantique des politiques de protection de la vie privée, la troisième catégorie concerne les approches visant à fournir des plateformes de représentation et le raisonnement sémantique sur les politiques de contrôle d'accès exprimées dans le langage d'ontologie OWL/RDF.

Bien que ces approches offrent un niveau d'expressivité satisfaisant pour définir les politiques de protection de la vie privée, elles n'offrent pas d'indications ou d'outils d'ingénierie adéquats facilitant la mise en oeuvre de systèmes de protection de la vie privée dans les scénarios cibles. De plus, ces approches se limitent dans l'ensemble à utiliser des mécanismes de raisonnement monotones, basés sur la logique de description selon l'hypothèse du monde ouvert. Ces mécanismes sont valables uniquement dans certains cas spécifiques et ne

permettent pas de définir des politiques de gestion de la vie privée en utilisant la négation de faits.

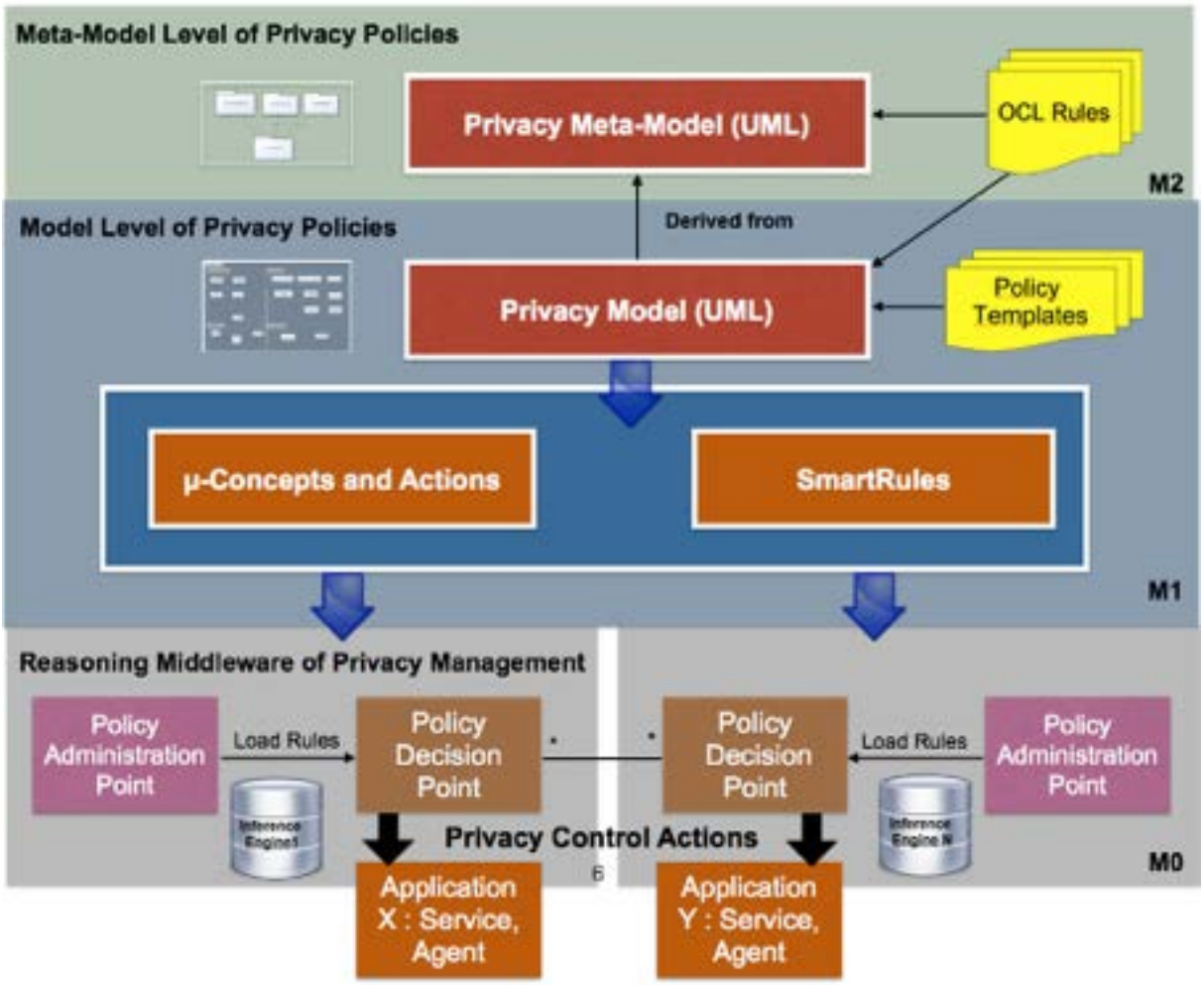
Pour répondre à l'ensemble des défis énumérés ci-dessus, nous proposons, dans cette thèse, un canevas sémantique intégrant un méta-modèle et des outils de raisonnement permettant de faciliter à tout concepteur de système ubiquitaire la mise en oeuvre de mécanismes de gestion des politiques de la vie privée qui soient efficaces, adaptatifs et sémantiquement interopérables. Dans l'approche proposée, l'objectif est d'exploiter la description conceptuelle et générique des connaissances contextuelles pour établir des règles génériques de protection de la vie privée afin de permettre à tout service ou agent de décider des contrôles à mettre en oeuvre pour protéger la vie privée de l'usager.

La description conceptuelle des entités de l'environnement et des politiques est fournie par le biais du méta-modèle proposé dans cette thèse. Le canevas proposé intègre aussi une architecture middleware générique qui offre des composants pour définir, administrer et contrôler l'application des politiques.

Pour ce faire, nous avons opté pour une approche hybride basée sur l'ingénierie dirigée par les modèles (Model Driven Engineering MDE) et un raisonnement à base d'ontologies et de règles d'inférence opérant selon l'hypothèse du monde fermé. Pour pallier aux limites des approches existantes, nous nous appuyons sur le standard « Model Driven Architecture » (MDA) de l'OMG pour définir un méta-modèle permettant de définir des politiques de protection de la vie privée qui soient intelligibles par des systèmes hétérogènes.

C'est ainsi que nous proposons un canevas sémantique pour la protection de la vie privée. L'architecture de ce canevas est fondée sur trois couches correspondant aux niveaux de modélisation de la protection de la vie privée : le niveau du méta-modèle, le niveau du

modèle et le niveau de la mise en œuvre du contrôle de la vie privée à travers une plate-forme



middleware pour le raisonnement. La figure 1 illustre l'architecture de ce canevas.

Figure 1 : Canevas sémantique

Le niveau du Méta-modèle offre un méta-modèle pour la description conceptuelle de la politique de protection de la vie privée. Ce méta-modèle est non seulement une représentation graphique ou formelle de concepts précieux pour obtenir le modèle de connaissances plus lisible pour les politiques de confidentialité, mais il fournit aussi, une approche beaucoup plus cohérente, conforme à l'approche dirigée par les modèles, pour la description de bon sens de ce que doit être la protection de la vie privée.

Le méta-modèle proposé, caractérisé par un niveau d'abstraction et d'expressivité élevé, permet de définir des politiques de protection de la vie privée indépendamment du domaine

d'application, et où ces politiques peuvent être adaptées aux différents contextes. Il hérite des concepts et règles de deux méta-modèles standardisés par l'OMG, en l'occurrence, ODM (Ontology Definition Metamodel) et BPDM (Business Process Definition Metamodel). ODM permet d'utiliser tous les constructeurs du langage OWL pour représenter les ontologies dans la notation UML. BPDM permet, quant à lui, de modéliser des politiques de protection de la vie privée telles que des obligations en utilisant les opérateurs de contrôle de processus. La relation entre le méta-modèle, ODM et BPDM est illustrée à travers la figure 2.

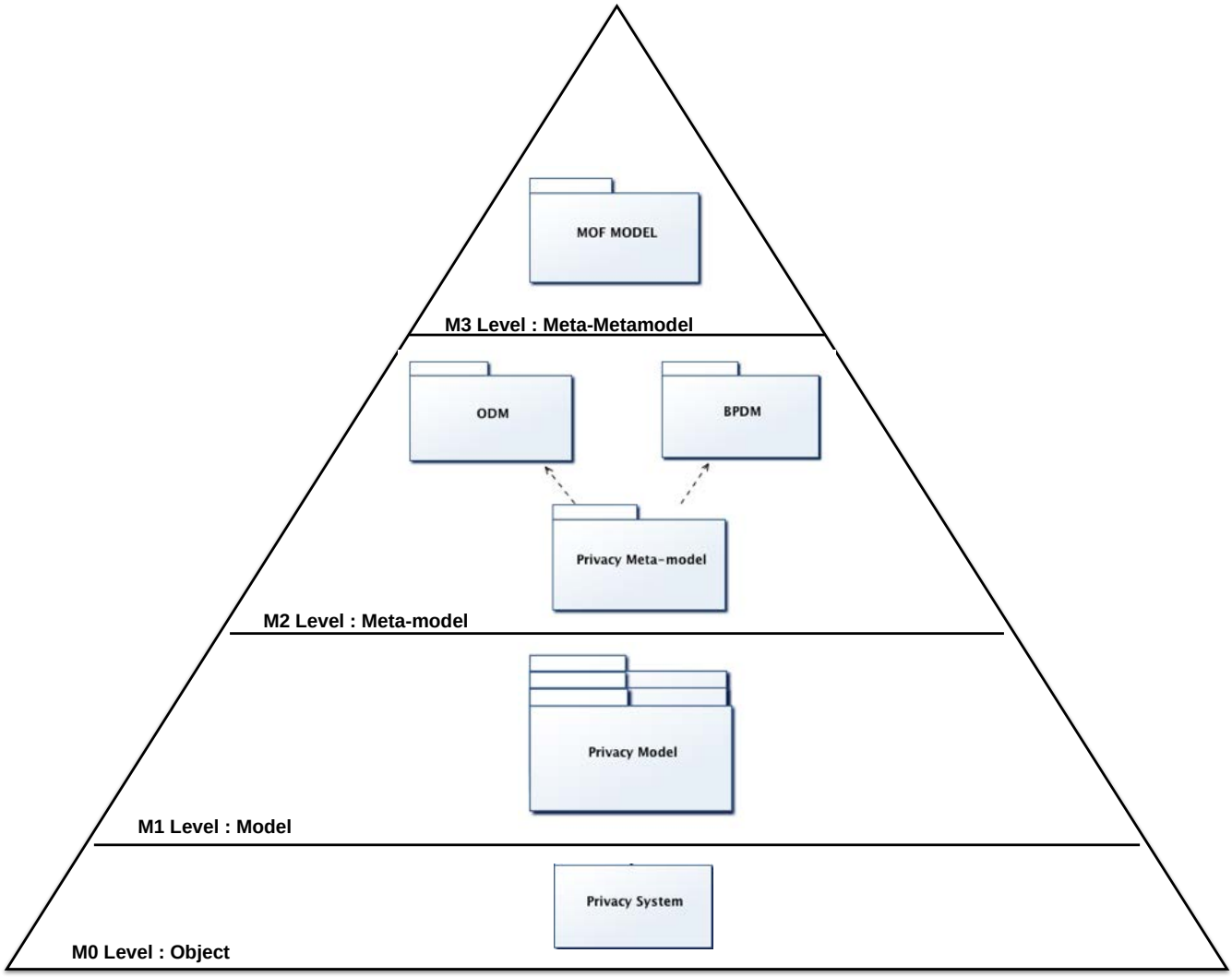


Figure 2 : Méta-modèle et MOF

Pour le méta-modele proposé, nous associons des modèles n-aires de politiques de protection de la vie privée et des contraintes du langage « Object Constraint Language (OCL) » pour la définition et la validation des politiques de protection de la vie privée. Nous proposons quatre

modèles de politiques qui couvrent ainsi les cas d'utilisation de la protection de la vie privée, à savoir : (i) les politiques de divulgation pour contrôler la façon dont les données personnelles sont divulguées, son propriétaire et à qui sera réalisée la transmission, (ii) les politiques de gestion de données pour contrôler la transmission, le traitement et le stockage des donnés personnelles par une partie tierce,
(iii) les politiques pouvant être appliquées aux capteurs pour contrôler les observations et finalement (iv) les politiques pour la protection de la divulgation des politiques de protection de la vie privée à une partie tierce.

Les politiques de protection de la vie privée peuvent être définies selon des modèles spécifiques. Une règle de protection de la vie privée est préfixée par le nom du modèle de la politique et composée de deux parties : antécédent et conséquent. Le méta-modèle propose des concepts de modélisation pour définir une politique en fonction de la forme générale de l'antécédent puis du conséquent. Cette règle peut être lue de manière informelle comme suit : si une opération sensible à la vie privée se produit sur une ressource privée et dans un contexte particulier alors appliquer décision de contrôle sur la vie privée qui peut être une action atomique réactive ou un processus pouvant comprendre les obligations. Nous proposons un modèle de politique de protection de la vie privée qui fournit une sorte de grammaire que nous avons utilisé pour la conception de notre méta-modèle pour la définition des politiques.

Les opérations sensibles à la vie privée concernent les opérations d'observation, de transmission, de traitement, de stockage et modification. Pour l'observation, l'utilisateur peut, tout simplement, lire des informations ou ressources dans un contexte très spécifique. Pour ce faire, des mesures nécessaires qui sont liées à l'observation sont « permettre » (ou autoriser) et « refuser » (ou interdire) l'observation.

La transmission de données sensibles et personnelles est considérée comme l'opération la plus délicate et critique dans la protection de la vie privée de l'utilisateur. Toute information envoyée aux particuliers, aux systèmes d'information ou à un site web peut être utilisée en toute circonstance, y compris, mais ne se limite pas à la diffusion, la reproduction, la transmission, la publication, la diffusion et la mise en place de contenu sur le web. Par

exemple, si un e-mail est utilisé pour transmettre cette information, les mécanismes de sécurité sont nécessaires pour le courrier électronique standard. Cela peut conduire à de graves abus, car il ouvre la voie pour que les données soient utilisées à des fins tout à fait différentes de son utilisation prévue. Ceci peut se produire pour un certain nombre de raisons. Les règles définissent qui peut utiliser les données et à quelles fins ou ceux qui contrôlent les données peuvent ne pas respecter les règles.

L'opération de traitement concerne, en général, l'utilisation de renseignements personnels comme entrée pour tout traitement. Un exemple typique est d'utiliser la liste des contacts de personne dans un système de diffusion. L'opération de modification est un type de traitement spécifique pouvant accorder à un tiers le droit limité de révéler des informations privées, mais il peut obliger le tiers à faire une transformation de ces données (telles que la transformation de l'âge à une catégorie d'âge).

Les utilisateurs n'ont plus le stockage physique de leur données. Cependant, l'opération de « stockage » permet à l'utilisateur de contrôler le stockage de ses données personnelles telles que les données médicales ou des données de localisation, dans le système. En d'autres termes, les détenteurs de données eux-mêmes sont responsables de divulguer l'information en autorisant ou interdisant le stockage des renseignements personnels.

Grâce aux actions de contrôle de la protection des renseignements personnels suivants, notre but est d'améliorer l'intégrité de l'information transmise par les opérations d'autorisation, d'anonymisation, d'obfuscation, de divulgation et des actions de restriction.

Pour l'action d'autorisation, l'individu peut autoriser ou interdire l'accès à ses données personnelles ou à ses ressources. Concernant l'anonymization, afin de prévenir les attaques de la vie privée, les données doivent être rendues anonymes bien avant qu'il soit divulguées. L'anonymisation correspond à une technologie qui convertit les données en texte clair dans une forme lisible et irréversible non humain. Cette technologie ne se limite pas aux techniques du hachage et celles du cryptage, dans lequel la clé de décryptage a été mis au rebut.

Dans notre thèse, nous considérons seulement l'anonymisation dans les environnements d'intelligence ambiante et les interactions des utilisateurs. En d'autres termes, l'anonymisation des données permet le transfert d'informations à travers une frontière, comme la transmission des données entre deux agents ou entre les utilisateurs, tout en réduisant le risque de divulgation involontaire d'une manière qui permet l'évaluation et l'analyse post-anonymisation. Nous notons qu'il existe plusieurs propositions de l'algorithme de k-anonymization dans la littérature. Les méthodes d'anonymisation devraient prendre en compte les modèles de protection de la vie privée La généralisation et la perturbation sont les deux approches d'anonymisation populaires pour les données personnelles. La généralisation est utilisée pour la collecte de données préservant la vie privée. Il existe plusieurs mesures de protection de la vie privée probabilistes basées sur une attaque de distribution et l'utilisent pour définir le problème de l'intimé de trouver un ensemble de données anonymes optimales.

La perturbation utilise des techniques aléatoires pour masquer les données pour préserver la confidentialité des données sensibles. Cette méthodologie tente de masquer les données sensibles en modifiant aléatoirement les valeurs de données en utilisant souvent le bruit additif.

En général, pour des raisons de confidentialité et de conservation, l'obfuscation vise à rendre les données personnelles des utilisateurs plus difficiles à comprendre ou à lire. L'obfuscation de données est une forme de masquage des données dans laquelle les données sont volontairement brouillées pour empêcher l'accès non autorisé aux matières sensibles. Cette forme de cryptage en résulte des données inintelligibles ou confus. Les techniques données de l'obfuscation faussent les données afin de cacher l'information. De nombreuses techniques d'obfuscation de données ont été proposées et mises en œuvre pour préserver la vie privée des applications de données. Les personnes qui utilisent l'obfuscation devraient être en mesure d'équilibrer leur niveau de confidentialité souhaité contre leur qualité souhaitée des données personnelles telles que son emplacement, le nom, l'âge, etc.

Dans notre thèse, nous nous focalisons sur l'auto-divulgation qui est le processus communication de l'information de soit une autre personne. En effet, les renseignements

personnels sont partagés avec les utilisateurs ou divulgués aux systèmes. En outre, nous nous concentrons sur le concept de divulgation minimale qui considère que la divulgation de données personnelles à des tiers doit être limité et ne se produit que sous certaines conditions. Par ailleurs, les restrictions doivent être remplies avant que l'accès aux données personnelles soit accordé. Si une seule condition n'est pas satisfaite, l'accès ne devrait pas être accordé.

Le langage de définition de contraintes OCL (Object Constraint Language) est utilisé pour définir des contraintes sur la structure des politiques de gestion de la vie privée ou sur l'architecture des systèmes de gestion des politiques. Selon l'approche MDA, les contraintes définies dans le méta-modèle sont valables dans tous les modèles spécifiés au niveau M1, ainsi que dans les systèmes implémentés à partir de ces derniers au niveau M0. Le méta-modèle proposé définit dans ce cas un cadre conceptuel pour établir des modèles de règles génériques et décidables permettant de prendre des décisions de contrôle cohérentes pour la protection de la vie privée. Les concepts de ce méta-modèle sont illustrés par les trois modules : ContextManagement, UserManagement et CorePrivacyPolicy.

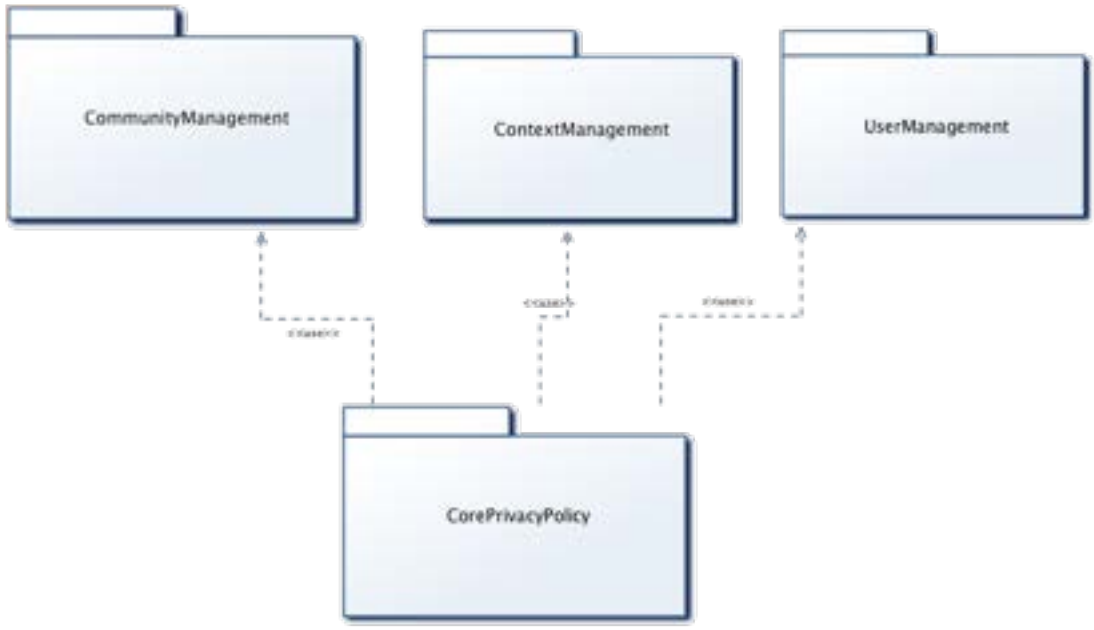


Figure 3 : Structure du méta-modèle

Ces modèles de règles sont mis en oeuvre au moyen du langage de règles SmartRules que nous avons développé dans le cadre du projet européen SembySem. Ce langage permet de répondre aux limites d'utilisation du langage SWRL. Il permet définir des règles basées sur la négation de faits et d'effectuer des inférences en chainage avant en utilisant des algorithmes de type Rete. Ces derniers, qui sont par nature décidable, sont particulièrement bien adaptés pour gérer de grandes bases de connaissances ainsi que les modèles de règles proposés dans notre canevas sémantique. Par ailleurs, l'utilisation des variables de concepts dans les règles SmartRules permet de mettre en oeuvre un contrôle adaptatif des politiques. Ce dernier est basé sur un raisonnement non-monotone et une représentation des instances de concepts selon la supposition du nom unique.

Le middleware que nous proposons aux concepteurs d'applications ambiantes permet de mettre en place un module de raisonnement. Cette approche présente une similarité avec la base XACML de système de contrôle d'accès où la décision d'autorisation et contrôle de l'application sont délégués aux composants spécifiques, à savoir la décision de politique point (PDP) et la politique Enforcement Point (PEP). Le point de la gestion des politiques (PAP) est utilisé pour stocker la politique de contrôle d'accès. PAP élabore des politiques de sécurité et stocke ces politiques dans le référentiel approprié. PDP évalue la politique applicable et rend une décision d'autorisation. Dans notre approche, le PDP et PEP sont encapsulés dans une entité «PDP», qui est un moteur d'inférence fonctionnant règles de production. Le PAP stocke les règles de la politique, exprimée dans la langue SmartRules, qui sont établies à partir des modèles de politiques.

Le canevas proposé a été validé à travers des scénarios typiques mettant en oeuvre des services d'assistance ambiante sensibles à la vie privée de personnes dépendantes. La mise oeuvre de chaque service selon le canevas proposé inclut le développement des modèles CIM (Computational Independent Model), PIM (Platform Independent Model) et PSM (Platform Specific Model).