Privacy model for federated identity architectures

Uciel Fragoso Rodriguez

Thèse de doctorat de l'INSTITUT NATIONAL DES TELECOMMUNICATIONS dans le cadre de l'école doctorale S&I en co-accréditation avec l'UNIVERSITE D'EVRY-VAL D'ESSONNE et en cotutelle avec ITAM, Mexique

Spécialité : Informatique

Par

M. Uciel FRAGOSO-RODRIGUEZ

Thèse présentée pour l'obtention du grade de Docteur de l'INSTITUT NATIONAL DES TELECOMMUNICATIONS

Modèle de Respect de la Vie Privée dans une Architecture d'Identité Fédérée

Soutenue le 16 décembre 2009 devant le jury composé de :

Mme. Isabelle CHRISTMENT	Rapporteur	LORIA, France
M. Francisco GARCIA-UGALDE	Rapporteur	UNAM, Mexique
Mme. Samia BOUZEFRANE	Examinatrice	CNAM, France
M. Ahmed SERHROUCHNI	Examinateur	TELECOM Paristech, France
Mme. Maryline LAURENT	Directrice	TELECOM SudParis, France
M. Jose INCERA-DIEGUEZ	Directeur	ITAM, Mexique

Thèse n° 2009TELE0026

Acknowledgments

I would like to acknowledge TELECOM & Management SudParis, the Université d´Evry and the ITAM for giving me the opportunity of making this doctoral thesis.

Also my acknowledgments to Isabelle Christment, Samia Bouzefrane, Francisco Garcia Ugalde and Ahmed Serhrouchni for being part of the jury committee; and specially to my advisors Maryline Laurent and Jose Incera for their advices and support during the development of the thesis.

Special acknowledgments to my family (Celia, Cintia and Manuel) for their encouragement and love, and also to my parents and all my family and friends from Mexico.

Résumé

L'Internet a provoqué une augmentation exponentielle du nombre de transactions en ligne. De même, il a impacté la façon dont les utilisateurs interagissent entre eux et avec d'autres organisations. Une grande partie des services de la vie quotidienne sont accédés de façon numérique. Les exigeances concernant l'offre de services en ligne sont de plus en plus fortes en termes de vélocité, disponibilité, mobilité et sécurité. Les entités qui fournissent ces services sont appelées Fournisseurs de Services (FS). Les FS doivent assurer l'intégrité et la confidentialité de l'information échangée dans les transactions en ligne. Quand il s'agit d'information personnelle, les FS devraient offrir garantir aux utilisateurs que celle-ci n'est pas compromise.

Pour chaque service fourni par les FS, les utilisateurs ont besoin d'une identité numérique. L'existence de multiples identités numériques représente une situation gênante aussi bien pour les utilisateurs que pour les FS. Pour les utilisateurs, il est compliqué de gérer beaucoup d'identités en accédant à plusieurs services en ligne; ils en font souvent l'amère expérience. En ce qui concerne les FS, chacun doit avoir son propre Système de Gestion d'Identités (SGI) pour gérer le cycle de vie des identités numériques, ce qui complique la collaboration quand il s'agit de fournir des services combinés. Prenons pour exemple le service de télévision mobile pour lequel deux FS collaborent : l'entreprise de TV qui fournit le contenu et l'entreprise mobile qui le distribue. Leurs deux SGI doivent pouvoir interférer pour lier les identités et traiter l'utilisateur comme une seule et même identité.

Une solution appropriée pour gérer de multiples identités dans un environnement distribué et collaboratif consiste à mettre en place une Architecture d'Identité Fédérée (AIF). Une AIF permet d'intégrer un ensemble d'organisations au travers d'accords commerciaux et d'une plateforme de technologie commune pour accéder aux services d'identité. « Single Sign On, SSO », fédération d'identités et échange d'attributs liés à l'identité en sont des exemples. Avant d'expliquer les fonctionnalités et la structure d'une AIF, il convient d'introduire certains termes comme l'identité numérique, les modèles de SGI et la relation de confiance.

Identité numérique

Une identité numérique peut être définie comme un ensemble de données numériques qui représentent de façon unique une entité dans un domaine d'application. Dans ce contexte, une entité peut être une personne, un ensemble de personnes, une organisation, un processus ou un dispositif, c'est-à-dire, tout objet capable de faire une transaction. Les éléments composant une identité numérique sont nommés *Attributs*, lesquels peuvent être assignés, intrinsèques à l'entité ou dérivés. Certains attributs distinguent de manière unique une identité numérique dans un contexte d'espace de noms: ils sont connus comme identifiants. Généralement, un identifiant est utilisé pour réaliser une authentification (c.-à.-d. valider l'identité). L'ensemble des éléments servant de preuve à l'authentification est appelé « credentials ». Un « credential » peut se présenter sous la forme d'un mot de passe, d'une réponse à un défi (quelque chose que l'on sait), d'une information fournie par une carte à puce ou un certificat numérique (quelque chose que l'on a), ou d'une information dérivée des caractéristiques de la personne, comme l'empreinte digitale, l'iris ou le timbre de la voix (ce que l'on est). La Figure 1 montre les relations qui existent entre les éléments composant une identité numérique.

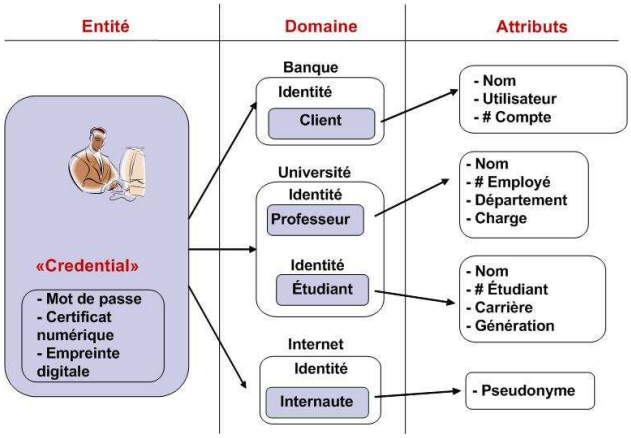


Figure 1. Relation entre les éléments composant une identité numérique

Dans la figure 1, l'entité a plusieurs identités numériques nécessaires pour interagir avec différents domaines d'application. Un domaine d'application est un contexte où une même identité numérique est valide, par exemple, une entreprise, un hôpital, un club sportif, une université ou l'Internet. On constate qu'une entité peut avoir plusieurs identités dans un même domaine d'application. Par exemple, à l'université, un professeur peut avoir simultanément l'identité de professeur et celle d'un étudiant dans le cas où il suit un cours en formation continue. Chaque identité est composée d'un ou plusieurs attributs et un « credential » associé.

Système de Gestion d'Identités

On définit un système de Gestion d'identités (SGI) comme l'ensemble des processus permettant de gérer le cycle de vie d'une identité numérique, c'est-à-dire, sa création, sa manipulation et sa fin de vie. Le SGI s'occupe aussi des composants opérationnels, lesquels gèrent les différents aspects de la sécurité, c'est-à-dire, le processus d'authentification, le contrôle d'accès et l'audit. La figure 2 montre les composants du cycle de vie et les composants opérationnels d'un SGI.

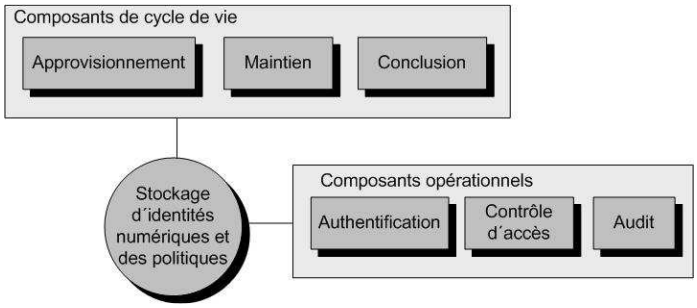


Figure 2. Composants et fonctionnalités d'un SGI

Modèles de Système de Gestion d'Identités

Les SGI ont beaucoup évolué. A ces débuts, chaque département dans une organisation gérait ses propres identités sans aucune intéraction facile possible (modèle isolé). Plus tard, des solutions centralisées ont été implémentées afin de n'avoir plus qu'à gérer qu'une seule identité de l'utilisateur (modèle centralisé). Aujourd'hui, il est nécessaire de gérer plusieurs identités dans un environnement distribué et collaboratif ; à cette fin, deux modèles ont été proposés : le modèle fédéré et le modèle centré sur l'utilisateur. Les paragraphes suivants expliquent avec un

certain niveau de détail ces modèles. Pour mieux comprendre la description de chacun des modèles, il est nécessaire de définit les concepts suivants :

Utilisateur.- Entité qui est représentée par une identité numérique et capable de réaliser une transaction.

Fournisseur de Service (FS).- Entité qui fournit un service aux utilisateurs ; il s'agit usuellement d'un site Web ou d'un service Web « Web Service ».

Fournisseur d'identité (FI).- Entité qui gère l'identité numérique de l'utilisateur et qui exécute le processus d'authentification.

Modèle isolé

Dans ce modèle, chaque FS a la responsabilité de gérer l'identité de chacun de ses utilisateurs. Le FS déploie sa propre SGI en prenant en compte la complexité et les fonctionnalités définies par l'organisation. Il s'avère très difficile, pour les FS, d'intégrer ces SGI afin de fournir des services coordonnés. De même, ce modèle devient assez lourd pour l'utilisateur lorsque le nombre d'identités à gérer augmente. En ce qui concerne le respect de la vie privée - ce qui est l'objet principal de ces travaux- le FS a le contrôle total de l'information des utilisateurs. Ceux-ci ont peu ou aucun contrôle sur les données personnelles gardées par le FS. La figure 3 montre les interactions entre un utilisateur et les FS pour le processus d'authentification et la fourniture de services.

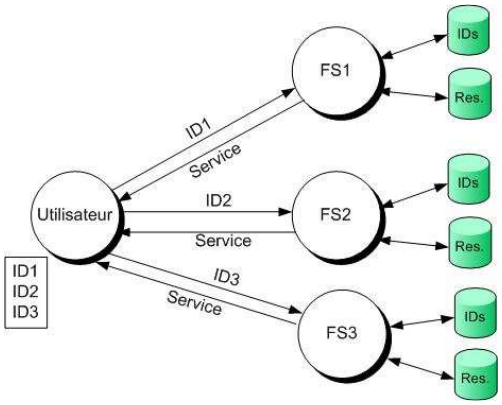


Figure 3. Modèle isolé

Modèle centralisé

Ce modèle repose sur le stockage unique des identités numériques. L'utilisateur peut s'authentifier avec tous les FS en utilisant la même identité. Il est donc assez simple d'implémenter la fonctionnalité de « Single Sign On » où l'utilisateur, une fois authentifié par le FI, peut accéder à plusieurs FS, et ce, sans authentification supplémentaire. Le modèle centralisé ne nécessite aucune expertise préalable de l'utilisateur du fait de la grande simplicité d'accès à plusieurs services avec une seule identité. Cependant, ce modèle ne manque pas d'inconvénients, parmi lesquels on peut souligner sa vulnérabilité du fait que le stockage centralisé d'identités représente un point unique de défaillance. Il ne passe pas à l'échelle lorsque le nombre d'identités devient très grand. La figure 4 montre comment un utilisateur peut utiliser la même identité pour accéder à plusieurs FS.

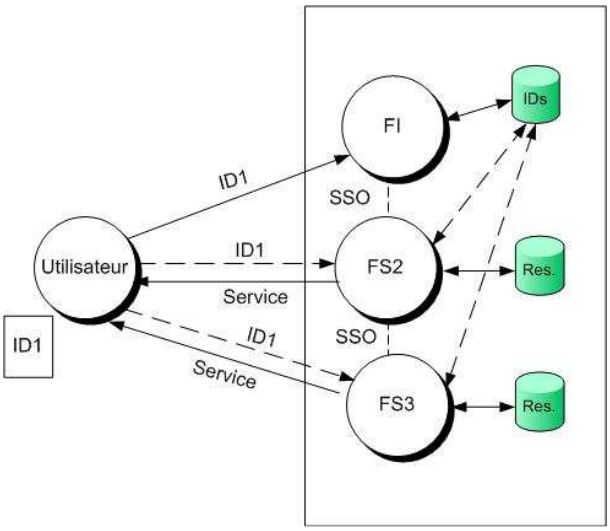


Figure 4. Modèle centralisé

Modèle fédéré

Dans le modèle fédéré, les identités sauvegardées dans les différents FS sont liées au travers de pseudonymes. Les entités qui composent la fédération forment un Cercle de Confiance (CC) en établissant des relations de confiance avec des accords commerciaux et une plateforme technologique commune. Le modèle fédéré définit des services d'identité tels que le SSO, la fédération d'identités et l'échange d'attributs. Tout comme dans le modèle précédent, une fois l'utilisateur authentifié avec le FI, il peut avoir accès aux services fournis par d'autres FS sans authentification supplémentaire. Or, comme le stockage est distribué, il n'a pas de point unique de défaillance, ni de limitation en principe quant au facteur d'échelle. Par contre, dans ce modèle, l'utilisateur n'a pas de contrôle sur ses données personnelles et il n'a aucune garantie quant au respect de sa vie privée. La Figure 5 montre les éléments d'un système fédéré et les relations entre eux.

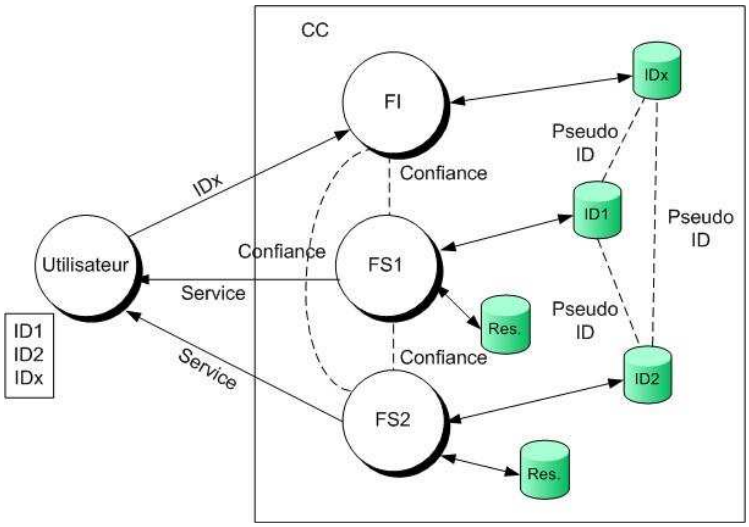


Figure 5. Modèle fédéré

Modèle centré utilisateur

Ce modèle a été proposé dans le but de donner aux utilisateurs plus de contrôle sur ses données personnelles. Il peut en effet sélectionner le FI qui lui convient et choisir l'identité à utiliser pour accéder aux différents FS. Les FS n'établissent pas de relation de confiance entre

eux pour fournir des services à l'utilisateur. Ce modèle permet à l'utilisateur de mettre en place son propre FI dans son ordinateur ou son portable. Parmi les défauts de ce modèle, on trouve la difficulté d'intégrer les FS pour fournir des services coordonnés, et le fait que le dispositif de l'utilisateur devient un point unique de défaillance. La figure 6 montre les éléments d'un système centré utilisateur et les relations entre eux.

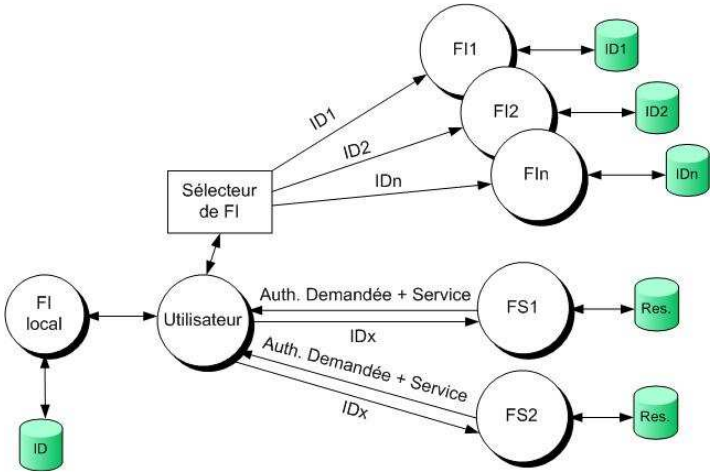


Figure 6. Modèle centré dans l'utilisateur

Modèles de confiance

Une des caractéristiques fondamentales d'un système de Gestion d'identités réside dans le niveau de confiance qui doit exister entre ses composants. Dans ce contexte, une relation de confiance peut être définie par la combinaison d'accords commerciaux et d'une plateforme technologique d'authentification. Les accords peuvent être établis de façon directe ou indirecte selon qu'il existe des intermédiaires ou non. Quand les accords sont établis directement, on parle d'un Modèle de Confiance par Pair, alors que si la relation est établie par un tiers, le modèle de confiance est connu comme Modèle de Confiance par Intermédiaire. Le mécanisme d'authentification peut être direct ou indirect. L'authentification directe est réalisée par l'utilisation de clés cryptographiques symétriques, tandis que l'authentification indirecte est établie par l'utilisation d'un ou plusieurs intermédiaires de façon qu'un chemin d'authentification puisse être dérivé entre les deux entités.

Architecture d'Identité Fédérée

Comme il a été mentionné précédemment, une Architecture d'Identité Fédérée (AIF) est composée d'un ensemble d'organisations qui ont établi des relations de confiance entre elles afin d'échanger des données de manière sûre, tout en préservant l'intégrité et la confidentialité de l'information. Dans ces travaux, on s'intéresse aux données personnelles.

Le FI dans l'AIF gère l'information d'identité de l'utilisateur et fait le processus d'authentification. Il peut y avoir un ou plusieurs FI dans le Cercle de Confiance. L'AIF doit accomplir les fonctionnalités suivantes du point de vue des utilisateurs, fournisseurs d'identités et fournisseur de services :

- « Single Sign On, SSO ».- SSO permet aux utilisateurs de s'authentifier avec un FI et d'accéder aux services fournis par plusieurs FS sans avoir besoin de s'authentifier à nouveau.
- Fédération d'identités.- Il s'agit de l'union de deux identités numériques au travers d'un pseudonyme pour implémenter les services de SSO et d'échange d'attributs.
- Échange d'attributs.- Le FS peut demander des attributs additionnels au FI pour fournir des services personnalisés.

Fédération d'identités

Quand les identités sont fédérées, un identifiant est créé pour chaque couple de FI et FS dans le but de lier les deux identités. L'identifiant peut être dynamique, c'est-à-dire, il est créé à chaque nouvelle session de l'utilisateur ou il peut être fixe pendant une longue période de temps. Un identifiant de type pseudonyme permet de préserver l'identité réelle de l'utilisateur et de mieux respecter sa vie privée. L'accord commercial établit la manière dont les identités sont fédérées, c'est-à-dire, la structure du pseudonyme, si l'identificateur est permanent ou dynamique, et les attributs échangés. La figure 7 montre la fédération d'identités qui utilise un pseudonyme fixe. Le tableau des identités de FI_1 montre comment l'identité ID1 est associée à l'identifiant aléatoire 65ER4589 quand elle est fédérée avec l'identité ID2 de FS1. Simultanément, le tableau des identités de FS_1 montre la relation existante entre l'identité locale ID2 et l'identificateur 65ER4589. Le pseudonyme a une couverture locale : le FI et le FS ne connaissent que le compte local et le pseudonyme. Quand deux entités ont besoin d'interagir pour échanger des informations d'identité, ils utilisent le pseudonyme pour la référencer.

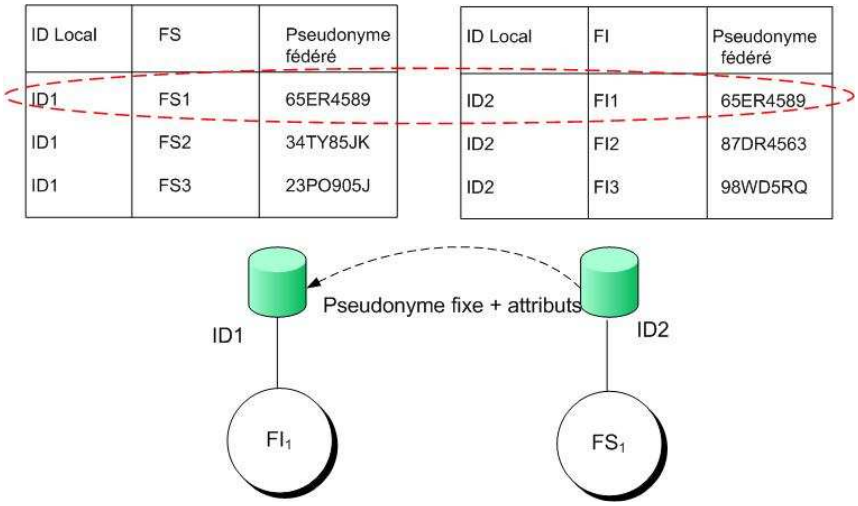


Figure 7. Fédération d'identités avec pseudonymes fixes

Respect de la vie privée

Le respect de la vie privée est un concept difficile à saisir parce que sa signification dépend du contexte, y compris du lieu géographique. Cela dit, le respect de la vie privée est identifié comme un droit fondamental dans tous les traités internationaux, les accords importants sur les droits de l'homme et dans les constitutions de la plupart des pays du monde. De ce fait, la majorité des pays ont développé des cadres juridiques concernant le respect de la vie privée.

Or, les technologies de l'information (TI) se sont développées très rapidement depuis les années 70, sans tenir compte, dans la plupart des cas, des droits de base du respect de la vie privée tels que définis dans les traités internationaux initiaux. Dans le même temps, ces traités se sont avérés insuffisants pour prendre en compte toutes les possibilités d'interaction des TI. Cette situation a nécessité la définition d'un nouveau cadre légal international, régional ainsi que local pour adresser le respect de la vie privée, comme la protection des données personnelles numériques contre la collection, la préservation et la diffusion inappropriées.

Depuis le dernier siècle, deux cadres légaux de respect de la vie privée ont vu le jour ; le premier est un modèle général centré sur les données à protéger et le dernier est un modèle sectoriel centré sur la personne et le contexte de traitement de données personnelles. Le modèle général vient du règlement européen dérivé des deux initiatives internationales les plus importantes : celle émise par l'Organisation pour la Coopération et le Développement Économique (OCDE) et la seconde émise par le Conseil de l'Europe. Le modèle légal sectoriel a été développé principalement aux Etats-Unis ; il se focalise sur des secteurs économiques avec

des autorégulations du marché et le gouvernement participe uniquement en tant que surveillant de secteur. Beaucoup d'autres lois et de recommandations sont apparues dans différentes régions géographiques, mais la plupart d'entre elles se basent sur les principes indiqués par les initiatives de l'OCDE et du Conseil de l'Europe.

Lignes directrices de l'OCDE

L'OCDE est une organisation constituée de 30 membres qui partagent le même engagement vis-à-vis de la démocratie et d'une économie de marché. L'OCDE est la première organisation internationale à avoir essayé d'unifier les initiatives relatives à la protection des données personnelles. En 1980, elle a publié les « Lignes directrices sur la protection de la vie privée et les flux transfrontières de données de caractère personnel ».

On peut condenser les directives de l'OCDE par les huit principes suivants, qui sont la base pour le développement de la majorité des initiatives mentionnées:

1. **Limitation en matière de collecte**.- « Il convient de fixer des limites sur la collecte de données à caractère personnel et toute information de ce type doive être obtenue par des moyens licites et loyaux et ceci après avoir informé la personne concernée ou avec son consentement ».
2. **Qualité des données**.- « Les données à caractère personnel doivent être congruents aux finalités en vue desquelles elles seront utilisées et, dans la mesure où ces finalités l'exigent, elles doivent être exactes, complètes et tenues à jour ».
3. **Spécification des finalités**.- « Les finalités en vue desquelles les données à caractère personnel sont collectées doivent être déterminées au plus tard au moment de la collecte. Elles ne doivent être utilisées par la suite que pour atteindre ces finalités ou d'autres qui ne sont pas incompatibles avec les précédentes et comme sont spécifiés à chaque occasion de changement de finalité».
4. **Limitation de l'utilisation**.- « Les données à caractère personnel ne doivent pas être divulguées, ni fournies, ni utilisées à d'autres fins que celles spécifiées conformément au principe de la spécification des finalités ».
5. **Garanties de sécurité**.- « Il convient de protéger les données à caractère personnel, moyennant des garanties de sécurité raisonnables, contre des risques tels que la perte des données ou leur accès, destruction, utilisation, ou divulgation non autorisés ».
6. **Transparence**.- « Il convient d'assurer, de façon générale, la transparence des progrès, pratiques et politiques, ayant trait aux données à caractère personnel. Il doit être possible de se procurer aisément les moyens de déterminer l'existence et la nature des données à caractère personnel, et les finalités principales de leur utilisation, ainsi que l'identité du dépositaire du fichier et le siège habituel de ses activités ».
7. **Participation individuelle**.- « Toute personne doit avoir le droit d'obtenir du dépositaire du fichier, ou par d'autres voies, confirmation du fait que le dépositaire du fichier détient ou pas des données la concernant ».
8. **Responsabilité**.- Tout dépositaire de fichier doit être responsable du respect des mesures donnant effet aux principes énoncés ci-dessus.

Les lignes directrices concernant le respect de la vie privée représentent le consensus international sur des conseils généraux au sujet de la collecte et de la gestion de l'information personnelle. Les principes de respect de la vie privée définis sont caractérisés par leur clarté et leur flexibilité pour s'adapter aux évolutions technologiques. Les principes s'appliquent aux niveaux nationaux et internationaux.

Directives du Conseil de l'Europe

En 1981, le Conseil de l'Europe a adopté la « Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel » connues sous le nom de « convention 108 du Conseil de l'Europe ». Cette convention est le premier instrument international obligatoire qui protège l'individu contre des abus. Elle peut accompagner la collection et le traitement des données personnelles et cherche à régler en même temps le flux transfrontière des données personnelles.

En 1995, l'Union Européenne a adopté la directive (95/46EC) sur la « protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données » connue sous le nom de « directive de protection des données d'UE ». Cette proposition était la réponse à la perception des insuffisances des recommandations concernant le respect de la vie privée de l'OCDE et la convention 108 du Conseil de l'Europe. D'après les termes des directives d'UE, chaque pays de l'UE doit avoir un commissaire ou une agence de protection des données qui impose les règles. On s'attend à ce que les pays avec lesquels l'Europe fait des affaires, fournissent un niveau de régulation identique.

Initiatives des Etats-Unis

L'approche américaine liée au respect de la vie privée est conduite par des intérêts commerciaux et emploie une approche de secteur qui se fonde sur un mélange des législations, des règlements, et d'autorégulation. Depuis l'adoption de la directive en 1998 par l'UE, le Ministère du Commerce des États-Unis a commencé une négociation intense avec la Commission Européenne afin de résoudre leurs divergences de politiques en matière de respect de la vie privée. Le résultat de ces négociations s'est concrétisé par le développement du cadre *Safe Harbor* qui exige des organismes des États-Unis d'assurer une protection équivalente en matière de vie privée, à celle définie par les directives de l'Union Européenne.

Un exemple d'initiative de respect de la vie privée focalisé sur les secteurs est la Portabilité d'Assurance Médicale et la Loi de Responsabilité de 1996 établie par le département des services de la santé et des affaires sociales des Etats-Unis. Cette loi adresse la divulgation et l'utilisation des informations liées à la santé des citoyens, le but principal étant d'assurer que l'information sur l'état de santé individuelle soit correctement protégée tandis que les soins de santé de qualité soient fournis. Un autre cadre juridique spécifique est publié par la Commission Commerciale Fédérale des États-Unis concernant la protection des données personnelles des consommateurs qui traitent des entreprises, aussi bien que des compagnies financières.

Technologies d'intrusion et d'amélioration du respect de la vie privée

On trouve aujourd'hui certaines technologies qui représentent un risque sérieux pour le respect de la vie privée; elles sont appelées Technologies d'Intrusion au Respect de la vie privée (TIR). Or, il existe aussi des contre-mesures technologiques connues sous le nom de Technologies d'Amélioration de Respect de la vie privée (TAR).

Technologies d'Intrusion au Respect de la vie privée

Les identités numériques sont sensibles aux menaces qui compromettent le respect de la vie privée des entités. Ces menaces peuvent être classifiées comme technologiques et non-technologiques ; ces dernières sont liées aux événements tels que le vol physique de l'information, sa perte dérivée d'une catastrophe naturelle ou l'ingénierie sociale. En ce qui concerne les menaces technologiques, nous nous concentrerons sur les technologies actuelles qui peuvent représenter un risque sérieux au respect de la vie privée des données personnelles. De telles technologies peuvent être classées selon leur domaine d'application comme suit :

1. **Systèmes d'identification**.- Les systèmes d'identification sont les systèmes et les technologies employés pour identifier une personne dans un domaine d'application.

Un élément très utilisé dans ces systèmes est la carte d'identification qui peut être employée pour des applications spécifiques. Les cartes peuvent stocker beaucoup de données personnelles qui peuvent être exposées ou abusées pendant le processus d'authentification. Deux technologies émergeantes ont été récemment incorporées aux cartes d'identification : la biométrie et l'identification par radiofréquence (RFID) ; elles sont employées pour faciliter le processus d'authentification. Les données biométriques sont très sensibles parce qu'elles sont fortement liées à la personne, tandis que la technologie RFID facilite la dissémination d'information sans avoir besoin d'un contact direct. Ces technologies peuvent représenter un risque sérieux au respect de la vie privée de données personnelle si elles ne sont pas protégées correctement.

2. **Surveillance des communications**.- La surveillance électronique est la technique d'interception des communications numériques. Elle est employée dans des cas spéciaux tels que le support technique ou la recherche criminalistique. Cependant, elle représente une violation du respect de la vie privée car la collecte des données personnelles a lieu sans la connaissance et le consentement de la personne concernée. De nombreuses technologies permettent de surveiller les communications. Quand les technologies de surveillance électronique sont combinées avec des systèmes telles que la téléphonie cellulaire, les systèmes de Wi-Fi ou le GPS (système de positionnement global), non seulement les informations personnelles peuvent être interceptés, mais aussi la localisation physique des utilisateurs peut être déterminée avec précision.
3. **Navigation de Web**.- L'une des applications les plus utilisées en support des transactions en ligne est le service Web, soutenu par le protocole HTTP. Ce protocole avec l'adresse IP du dispositif utilisé, permet de tracer l'utilisateur, de connaître sa localisation géographique ainsi que les caractéristiques de son navigateur Web. Bien que le fait de tracer l'utilisateur représente un risque de non respect de sa vie privée, la vraie menace pour lui est l'information véhiculée par ses *cookies*. Les informations collectées par des *cookies* peuvent être communiquées avec des informations personnelles telles que le nom, email, adresse, ainsi que des préférences personnelles.

Technologies d'Amélioration du Respect de la vie privée

Les Technologies d'Amélioration de Respect de la vie privée sont les technologies qui aident à la protéger. Elles peuvent procurer aux utilisateurs le contrôle pour décider de l'utilisation, révélation et distribution de données personnelles en ligne. Elles peuvent également aider les organismes à définir leurs propres politiques en matière de respect de la vie privée. Les TAR peuvent être classés en :

1. **Technologies du côté de l'utilisateur**.- Ces technologies sont déployées directement dans les dispositifs de l'utilisateur. L'une des technologies les plus utilisées est le chiffrement des données. Le mécanisme de chiffrement garantit dans une certaine mesure la confidentialité des données personnelles quand elles sont transmises ou sauvegardées. Il existe des technologies additionnelles qui peuvent être employées pour améliorer le respect de la vie privée, par exemple, les outils qui aident à contrôler les *cookies* (en les bloquant, en les sélectionnant, en les supprimant ou en les regardant). De telles fonctionnalités sont incorporées dans les versions récentes des navigateurs Web.
2. **Mécanismes des tiers**.- Ce type de mécanismes de respect de la vie privée sont déployés entre deux entités qui échangent des données personnelles. Quelques exemples de ces technologies sont : les pseudo-identités et l'anonymat, le contrôle d'accès et les langages de respect de la vie privée. Les mécanismes pour établir le pseudonyme et l'anonymat sont également connus comme protecteurs d'identité.

Un protecteur d'identité peut être vu comme l'intermédiaire entre l'utilisateur et le Fournisseur de Service ; il offre pour fonctionnalité de cacher la véritable identité de l'utilisateur pour des services où la connaissance de celle-ci n'est pas nécessaire.

3. **Langages de politique de respect de la vie privée**.- Les langages de politique en matière de respect de la vie privée peuvent aider dans plusieurs des étapes impliquées dans la gestion de ces politiques comme leur spécification, leur combinaison et leur mise en œuvre. Quelques langages de politique sont conçus pour aider les organismes à exprimer leurs politiques de respect de la vie privée et pour faciliter leur application, alors que d'autres sont conçus pour aider les utilisateurs à définir leurs préférences en matière de vie privée.

Échange d'attributs dans un système d'identité fédéré

Dans le système d'identité fédéré, l'échange de données personnelles (attributs) peut se produire entre n'importe quelle entité du cercle de confiance (utilisateurs, FS et FI). Quand l'utilisateur participe à l'échange d'attributs, il peut décider à quel destinataire les fournir et sous quelles conditions. Malheureusement, le scénario le plus répandu et celui qui pose un vrai risque de non respect de la vie privée a lieu quand le FS demande les attributs au FI afin de personnaliser et ainsi d'améliorer le service fourni. Dans ce cas, l'utilisateur ne peut pas contrôler ses données personnelles publiées par le FI.

La figure 8 montre un flux possible d'informations pendant le processus d'authentification et d'échange d'attributs entre le FI et le FS dans l'AIF. Dans ce scénario, le FI et le FS détiennent chacun des données personnelles liées à l'utilisateur, celles-ci sont fédérées par un pseudonyme. Le processus commence quand l'utilisateur s'authentifie avec le FI en suivant n'importe quelle méthode d'authentification définie par lui (1). Si l'authentification réussit, le FI donne à l'utilisateur un jeton d'identité (2) avec l'information d'authentification et un pseudonyme qui est employé pour accéder aux services fournis dans le cercle de confiance. L'utilisateur demande un service du FS et présente le jeton donné par le FI (3), le FS valide le jeton d'authentification avec le FI (4), puis traduit le pseudonyme en l'identité locale afin de pourvoir le service correspondant (5). Si le service demandé a besoin d'attributs supplémentaires qu'il n'a pas, ils sont demandés au FI (6), la pseudo-identification est employée pour référencer l'utilisateur. Les attributs sont envoyés au FS (7) et finalement, le service est fourni (8).

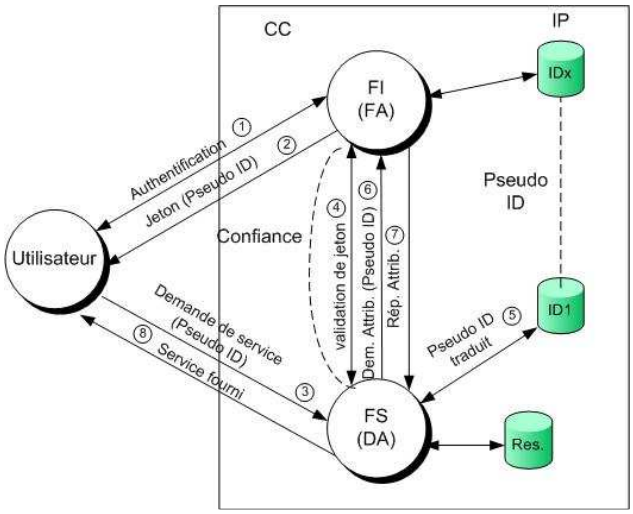


Figure 8. Une possible échange d'attributs dans une architecture d'identité fédérée

Mécanismes disponibles pour le respect de la vie privée dans une AIF

L'AIF offre quelques outils et directives pour établir des services d'identité plus sécurisés et qui tiennent compte, dans une certain mesure, du respect de la vie privée, principalement quand les attributs sont transmis du FI au FS. Parmi les outils et mécanismes disponibles, nous pouvons citer :

1. **Canal sécurisé**. - La communication entre tous les composants de l'AIF est effectuée d'une façon sécurisée à l'aide de canaux chiffrés et authentifiés, ce qui empêche n'importe quelle entité non autorisée d'intercepter les données personnelles.

2. **Message sécurisé**. - En plus de la sécurité du canal, les messages échangés par les entités composant un cercle de confiance sont chiffrés et signés électroniquement afin d'assurer la confidentialité et l'intégrité des données personnelles échangées.

3. **Pseudonyme**. - Les spécifications de l'AIF permettent l'attribution d'un nombre arbitraire de caractères par le FI ou le FS pour identifier un utilisateur (pseudo identificateur). Ils facilitent la fédération d'identité entre les différents comptes des utilisateurs dans le FI et le FS sans divulguer leur identité réelle.

4. **Anonymat**. - L'AIF peut utiliser un identifiant anonyme dans le but de partager des données avec le FS pour fournir des services. L'anonymat est employé quand le FS a un compte local pour l'utilisateur et qu'il n'est pas nécessaire de connaître sa véritable identité.

5. **Directives d'utilisation**. - Le protocole d'échange permet aux entités ayant besoin de transmettre des données personnelles (demandeur d'attribut et fournisseur d'attribut) d'intégrer les directives dans le message. Ces directives peuvent être employées pour spécifier l'utilisation prévue et l'utilisation autorisée des attributs à échanger. L'utilisation prévue et l'utilisation autorisée doivent coïncider afin de libérer les attributs demandés.

6. **Service d'interaction**. – L'AIF inclut un service d'interaction qui permet au FS de faire des interactions directes avec l'utilisateur, et ce, afin d'obtenir son consentement explicite pour une certaine utilisation de ses attributs.

Aspects manquant au respect de la vie privée dans une AIF

Malgré les mécanismes disponibles exposés ci-dessus, il existe plusieurs limitations concernant la manière dont le respect de la vie privée est géré dans les AIF :

- Les spécifications de l'AIF ne définissent pas comment les données personnelles (DP) sont rassemblées.
- Une fois que les DP sont collectés par une entité, l'utilisateur n'a plus de contrôle sur la façon dont ses DP pourrait être libérés à d'autres entités du cercle de confiance.
- Même si une option est prévue pour que le FS informe l'utilisateur quant au traitement de ses DP en termes de respect de la vie privée, il n'existe aucun mécanisme permettant de garantir sa bonne prise en compte.
- Quand le FS demande un attribut, le FI authentifie le demandeur et garantit que l'information transmise est chiffrée et signée électroniquement. Cependant, il n'existe aucun mécanisme de contrôle d'accès.
- Une fois les DP délivrés au FS, l'utilisateur ne sait pas à qui ses données personnelles sont libérées et dans quels buts. Il manque clairement un processus d'audit.

Corrélation des mécanismes de respect de la vie privée et des principes de respect de la vie privée.

Les mécanismes intrinsèques de l'AIF prévoient la mise en œuvre de quelques principes de respect de la vie privée spécifiés par le cadre de régulation ; cependant, quelques uns de ces aspects manquent et doivent être comblés par des fonctionnalités complémentaires. Le tableau 1 montre comment les mécanismes de respect de la vie privée proposés par l'AIF accomplissent dans une certaine mesure les principes de respect de la vie privée proposés par un cadre de régulation, et dans ce cas précis les directives de l'OCDE.

	Directives de l'OCDE							
	Limitation de collecte	Qualité des données	Spécifications de finalités	Limitation de l'utilisation	Garanties de sécurité	Transparence	Participation individuelle	Responsabilité
Mécanismes intrinsèques de respect de la vie privée de la l'AIF								
Canal sécurisé					■			
Message sécurisé					■			
Pseudonyme				■				
Anonymat				■				
Directives d'utilisation			■			■		
Service d'interaction				■				

Tableau 1. Mécanismes intrinsèques de respect de la vie privée de l'AIF et sa corrélation avec les directives de respect de la vie privée de l 'OCDE

Les communications chiffrées et authentifiées (canal et message sécurisés) entre les composants de l'AIF, sont les deux mécanismes les plus importants de « garanties de sécurité » afin d'assurer l'intégrité et la confidentialité dans l'échange de données personnelles. Le pseudonyme et les fonctionnalités d'anonymat aident à limiter la révélation de la véritable identité de l'utilisateur, ce qui satisfait dans une certaine mesure le principe de « limitation d'utilisation ». Les directives d'utilisation sont les messages échangés pour négocier l'utilisation prévue des données personnelles par le demandeur ainsi que l'utilisation permise par le fournisseur de service. Cette fonctionnalité peut être employée pour accomplir les principes de respect de la vie privée de « Spécifications de finalités » et de « Transparence ». Finalement, les possibilités de services d'interaction aident l'utilisateur à donner son accord quant à la révélation directe de ses informations personnelles, ce qui permet de supporter le principe de « Limitation de l'utilisation ».

On remarque, dans le tableau 1, que certains principes de respect de la vie privée ne sont pas pris en compte par les mécanismes intrinsèques de l'AIF. La section suivante propose un modèle de gestion de respect de la vie privée basé sur des politiques. Le modèle considère de nouveaux composants et fonctionnalités qui doivent être intégrés dans l'entité de l'AIF responsable du partage des données personnelles (le fournisseur d'attributs). Le modèle satisfait les principes de respect de la vie privée qui ne sont pas accomplis par les mécanismes intrinsèques de l'AIF.

Modèle de Gestion de Respect de la Vie Privée

Le modèle proposé de respect de la vie privée a pour objectif principal de compléter et de renforcer les aspects de respect de la vie privée de l'AIF afin de soutenir techniquement un cadre juridique de respect de la vie privée. L'idée fondamentale est de protéger l'information personnelle au moyen de politiques en matière de protection de la vie privée en prenant en considération des cadres légaux (externes et internes) et les préférences de l'utilisateur. Le modèle doit imposer les politiques en matière de respect de la vie privée et doit permettre d'auditer ses fonctionnalités principales qui sont nécessaires pour améliorer le niveau de respect de la vie privée dans un système de AIF ; le modèle fournit un ensemble de fonctionnalités qui permettent de :

- Définir des conditions de respect de la vie privée.
- Représenter les données personnelles dans une abstraction de données standard.
- Donner accès à l'information personnelle de l'utilisateur afin de la vérifier et de l'actualiser.
- Créer des politiques en matière de respect de la vie privée.
- Spécifier des préférences de respect de la vie privée des utilisateurs.
- Imposer des politiques.
- Auditer la conformité des politiques.

Les politiques en matière de respect de la vie privée sont définies d'une manière hiérarchique en prenant compte des régulations nationales et internationales, des politiques de l'organisation et des préférences de l'utilisateur. Les demandes de respect de la vie privée doivent pouvoir être traduites du langage naturel vers des ontologies, de sorte que les entités échangeant les informations personnelles puissent manipuler les mêmes définitions de syntaxe et de sémantique. Chaque activité de définition et d'application de politiques de respect de la vie privée doit être enregistrée afin qu'un système d'audit puisse les vérifier. Cette fonctionnalité est très importante pour que les utilisateurs sachent à tout moment comment leur information personnelle est traitée.

Dans le modèle proposé, nous nous basons sur les hypothèses suivantes : les organismes échangeant l'information personnelle appartiennent au même cercle de confiance et ont établi un accord commercial concernant les mécanismes de politique en matière de respect de la vie privée et un cadre de régulation compatible. L'architecture proposée est un modèle à 3 couches représenté dans la figure 9.

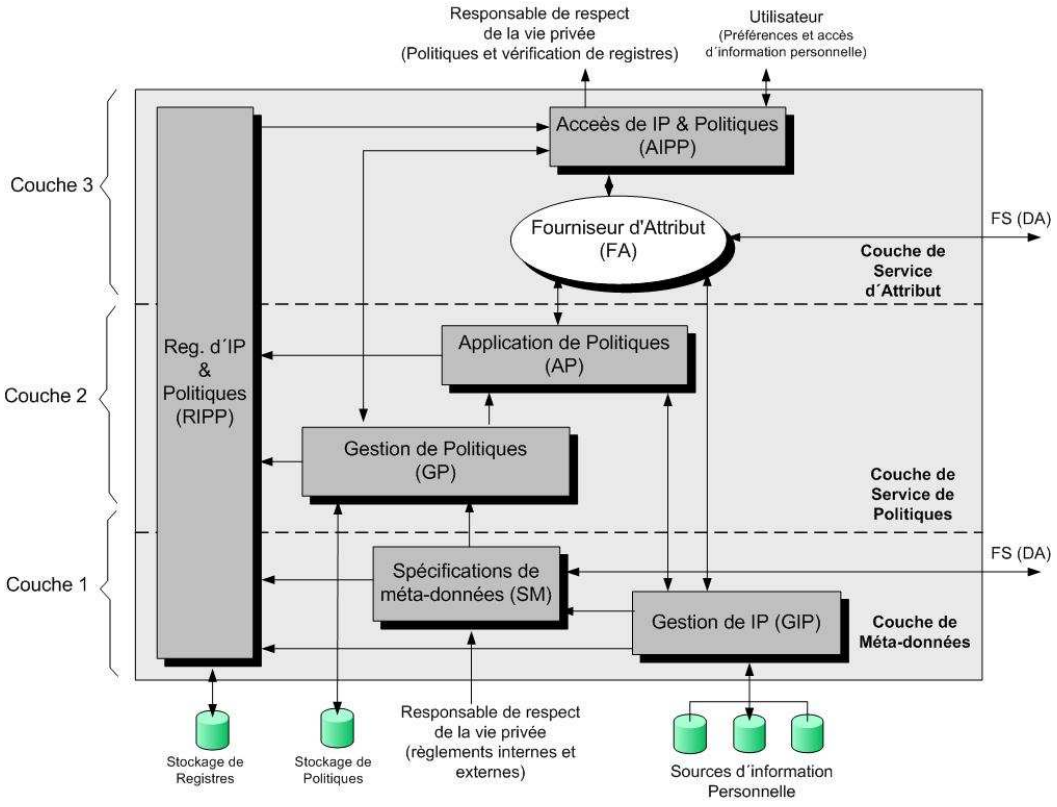
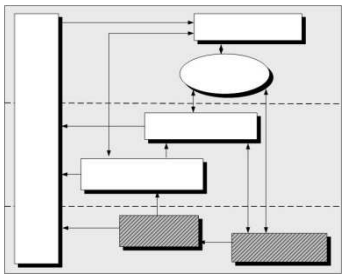


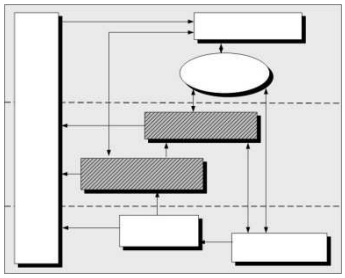
Figure 9. Architecture du modèle de respect de la vie privée

Couche de Méta-données



Elle fournit les fonctionnalités de base qui permettent de définir des conditions de respect de la vie privée et de gestion de l'information personnelle. Pour accomplir ces fonctionnalités, la couche est composée de deux modules : le module de Spécifications de Méta-données (SM) et le module de Gestion de l'Information Personnelle (GIP). Le module SM reçoit les spécifications de respect de la vie privée en termes de régulations externes et internes du responsable de respect de la vie privée et produit une ontologie qui corrèle l'information personnelle, les services et les spécifications de respect de la vie privée. Le module GIP fournit une interface normalisée aux points de données personnels hétérogènes.

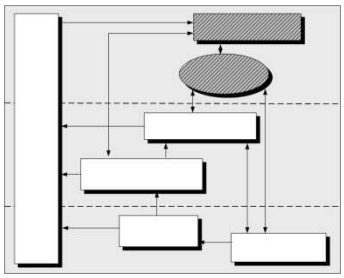
Couche de Service de Politiques



Elle fait la création et la gestion des politiques en matière de respect de la vie privée. Elle contient deux modules: le module de la gestion de politiques (GP) et le module de l'Application de politiques (AP). Le module GP crée les politiques de respect de la vie privée permettant aux utilisateurs de spécifier leurs préférences. Il indexe, stocke et recherche des politiques lorsqu'elles sont demandées par d'autres modules. Le module AP intercepte la demande d'attribut ; il confronte la demande à la politique de respect de la vie privée correspondante et puis

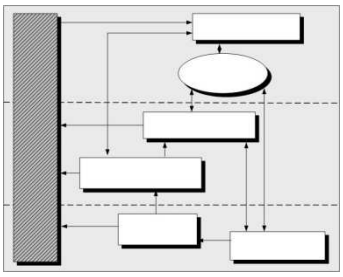
procède à l'évaluation. Le résultat de l'évaluation est renvoyé à l'entité de fournisseur d'attribut afin d'accepter ou refuser la délivrance de l'attribut.

Couche de Service d'Attribut



Elle permet d'accéder à l'information personnelle demandée par le fournisseur de service, ou directement par l'utilisateur ou par le Responsable de respect de la vie privée. Cette couche se compose de l'entité Fournisseur d'Attribut (FA) et du module d'Accès d'Information Personnelle et de Politiques (AIPP). L'entité FA est un composant original du FI. Elle est responsable de libérer les attributs demandés par un demandeur d'identité. Le module AIPP est une interface qui permet à l'utilisateur d'accéder à ses informations personnelles afin de les mettre à jour ou de vérifier sa conformité et son intégrité.

Module de Registre d'Information Personnelle et de Politiques



Toute l'activité qui se produit dans les deux couches inférieures est registrée par un module appelé Registre d'Informations Personnelle et des Politiques (RIPP). Cette information est à la disposition du Responsable de respect de la vie privée pour auditer le bon respect des finalités et pour permettre à l'utilisateur de contrôler à tout moment comment son information personnelle est gérée par le fournisseur d'attributs en termes de respect de la vie privée.

Modèle de respect de la vie privée et les principes de respect de la vie privée

Le modèle de respect de la vie privée précédemment présenté complète les caractéristiques intrinsèques de l'AIF afin d'augmenter la conformité de respect de la vie privée d'un cadre de régulation, dans notre cas, les directives de l'OCDE. Le tableau 2 est le complément du tableau 1; les caractéristiques intrinsèques de respect de la vie privée sont séparées de celles fournies par le modèle de respect de la vie privée proposé afin de mieux visualiser les apports du modèle.

Les mécanismes intrinsèques de respect de la vie privée de l'AIF se focalisent principalement pour garantir le respect de la vie privée pendant l'échange d'attributs, tandis que le modèle proposé augmente le respect de la vie privée en permettant aux utilisateurs et aux entités de spécifier et d'imposer des politiques appropriées. Le modèle souligne l'enregistrement de la plupart des événements afin de faciliter les processus d'audit.

On considère aussi une interface graphique qui rend possible l'accès aux données personnelles, politiques de respect de la vie privée et registres d'événements. Cela permet de satisfaire les principes de « limitation de collecte », « qualité de données », « transparence » et « participation individuelle ». Les caractéristiques de respect de la vie privée des organismes et les cadres juridiques externes définis à côté du FA (méta-données), permettent aux utilisateurs de savoir comment leurs informations personnelles vont être traitées par le FA, ce qui satisfait de cette façon le principe de « spécifications de finalités ». Le contrôle d'accès de politique assure la protection des informations personnelles en imposant les politiques et en révélant les informations personnelles uniquement aux entités autorisées et pour les finalités précédemment établies, donc il constitue l'une des mesures de sécurité les plus fondamentales, celles de la « limitation de l'utilisation » et la « garantie de sécurité ». Finalement, la fonctionnalité de registre proposée dans le modèle garantit le principe de « responsabilité » en permettant aux utilisateurs et à des auditeurs de savoir comment l´information personnelle est traitée pendant sa collection, sauvegarde et transmission. L'exécution de toutes les fonctionnalités du modèle

augmente sensiblement le niveau de respect de la vie privée de l'architecture fédérée d'identité pendant le processus du partage d'attributs.

	Directives de l'OCDE							
	Limitation de collecte	Qualité de données	Spécifications de finalités	Limitation de l'utilisation	Garanties de sécurité	Transparence	Participation individuelle	Responsabilité
Mécanismes intrinsèques de respect de la vie privée de la l'AIF								
Canal sécurisé					■			
Message sécurisé					■			
Pseudonyme				■				
Anonymat				■				
Directives d'utilisation			■			■		
Service d'interaction				■				
Les caractéristiques de respect de la vie privée ajoutées par le modèle								
Spécification de Méta-données			▤					▤
Gestion d'IP		▤					▤	▤
Interface d'IP et politiques	▤	▤				▤	▤	▤
Administration de politiques			▤	▤		▤		▤
Préférences de respect de la vie privée				▤				▤
Application de politiques				▤	▤			▤
Registre d'IP et politiques								▤

Tableau 2. Modèle de respect de la vie privée et sa corrélation avec les principes de respect de la vie privée

Mise en œuvre du modèle de respect de la vie privée dans un scénario du projet d'e-gouvernement Mexicain

Afin de décrire le processus de mise en œuvre du modèle de respect de la vie privée proposé, on propose un cas d'étude simplifié où les entités EgouvA et EgouvB représentent des organisations de gouvernement Mexicain. Elles appartiennent à un cercle de confiance où les identités numériques des utilisateurs sont fédérées. La fédération fournit des services d'identité comme le SSO et l'échange d´attributs. EgouvA joue le rôle de FI et EgouvB joue le rôle des FS. EgouvA contient le composant du fournisseur d'attributs (FA) qui libère l´information personnelle au demandeur d´attributs (DA). Les politiques en matière de respect de la vie privée sont associées aux informations personnelles et elles sont imposées aux attributs du côté de FI à chaque fois qu'ils sont demandés. La figure 10 montre les interactions entre l'utilisateur et les entités A et B quand l'utilisateur s'authentifie avec EgouvA et demande un service à EgouvB.

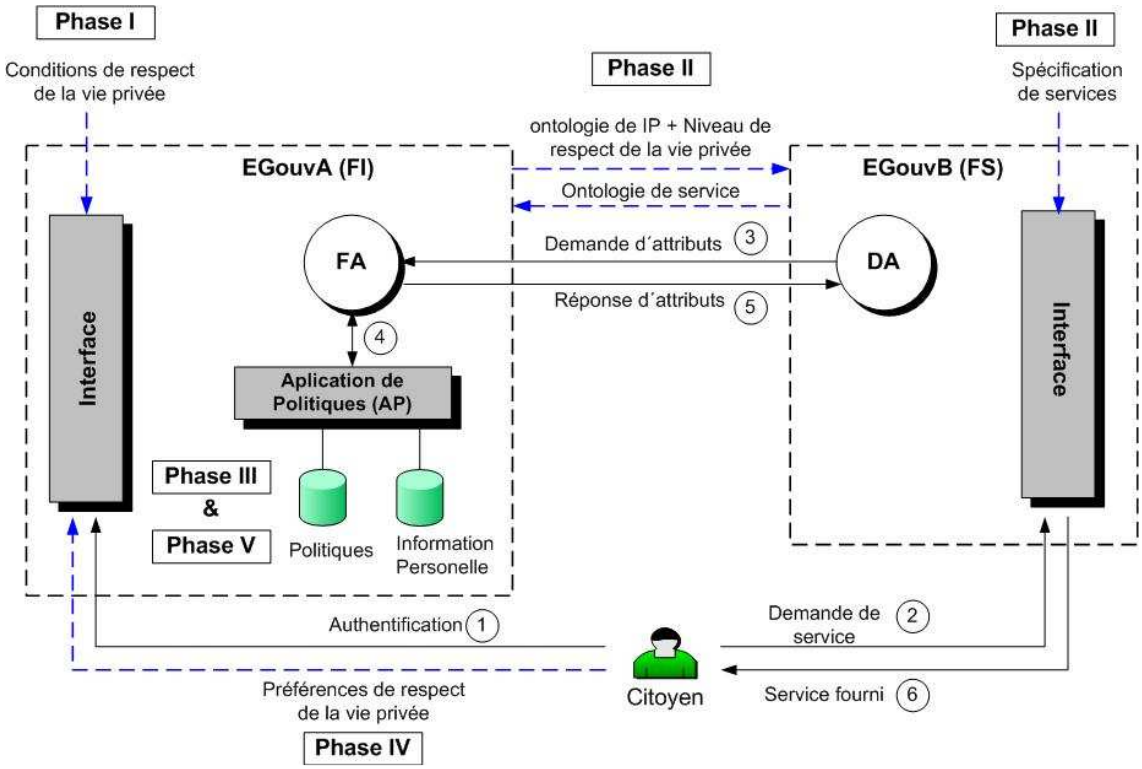


Figure 10. Scénario de mise en œuvre du modèle de respect de la vie privée

Une fois l'utilisateur authentifié avec EgouvA (1), il demande des services à EgouvB (2) ; si le service a besoin de données personnelles additionnelles, elles sont demandées (3) au fournisseur d'attribut (qui est dans ce cas, l'entité EgouvA). Le FA reçoit la demande qui contient : quel attribut est nécessaire, qui demande les données, et les paramètres de respect de la vie privée (la finalité et le temps de conservation) selon les politiques de respect de la vie privée qui sont associés aux attributs. Le FA analyse les paramètres de la requête et puis sollicite le module de l'application de politique (AP) afin d'évaluer la demande (4). Si l'évaluation s'avère positive, le FA envoie une réponse de confirmation positive au DA avec les attributs correspondants ; le cas échéant, une réponse négative est envoyée (5). Finalement, le service est fourni à l'utilisateur (6).

Dans ces travaux, on propose aussi une méthodologie du déploiement du modèle composée de cinq phases. Elle est représentée à la figure 10 et décrite avec un certain niveau de détails dans les paragraphes suivants.

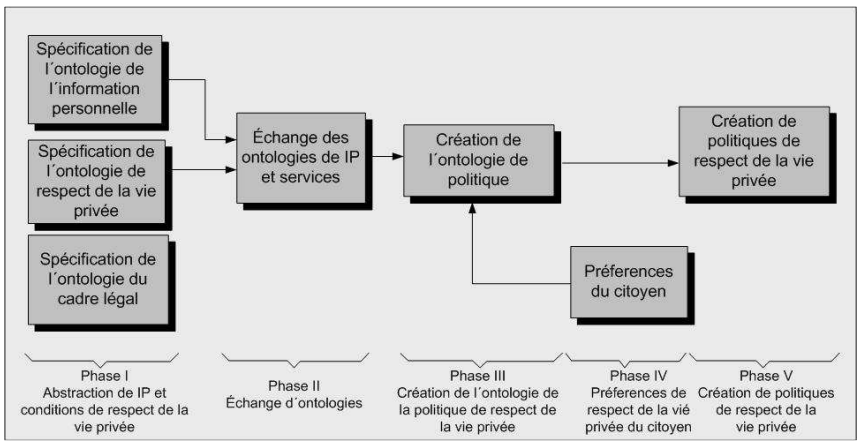


Figure 11. Phases pour déployer le modèle de respect de la vie privée

Phase I : Cette phase est effectuée par les entités responsables de la définition de l'ontologie de l´information personnelle de l'utilisateur. Pendant cette phase, les experts (responsables du respect de la vie privée) du cadre légal expriment les conditions externes de respect de la vie privée en langage naturel. Les spécifications associées à la vie privée sont converties en expression formelle par une méthodologie de développement d'ontologies. Les modules du modèle de respect de la vie privée impliqués dans cette phase sont : Spécifications de Méta-données et Gestion de IP (voir la figure 9).

Supposons que les spécifications des données personnelles et les niveaux de respect de la vie privée définis dans l´entité EgouvA sont exposés dans le tableau suivant:

Catégorie d'IP	***Attributs***	***Niveau de respect de la vie privée***
Données d'identification	*Nom et prénoms, Date de naissance, Nationalité*	*Publique*
Données académiques	*Titres professionnels, ID Professionnel*	*Modéré*
Données de santé	*Maladies, Allergies, État de santé, Traitements médicaux*	*Strict*

Les niveaux de respect de la vie privée sont définis comme:

***Niveau publique** : Les données personnelles sont partagées avec n'importe quelle entité dont les pratiques en matière de respect de la vie privée d'utilisation ne sont pas connues et ces données peuvent être employées pour n'importe quelle finalité. Des données personnelles peuvent être gardées indéfiniment ou aussi longtemps que la loi le permet.*

***Niveau modéré** : Les données personnelles sont partagées uniquement avec des entités connexes au service et avec le respect de la vie privée équivalente des pratiques et elles peuvent être employées pour personnaliser les services. Les données peuvent être gardées pendant toute la durée de la fourniture du service ou conformément à la loi.*

***Niveau strict :** Les données personnelles sont partagées uniquement avec l'entité fournissant le service et elles peuvent être employées seulement pour une finalité spécifique. Les données sont gardées seulement pendant la durée de la transaction du service ou selon les exigences de la loi.*

Les spécifications précédentes exprimées en langage naturel sont converties en ontologie et exprimé dans une représentation ORM (Object-Role Modeling):

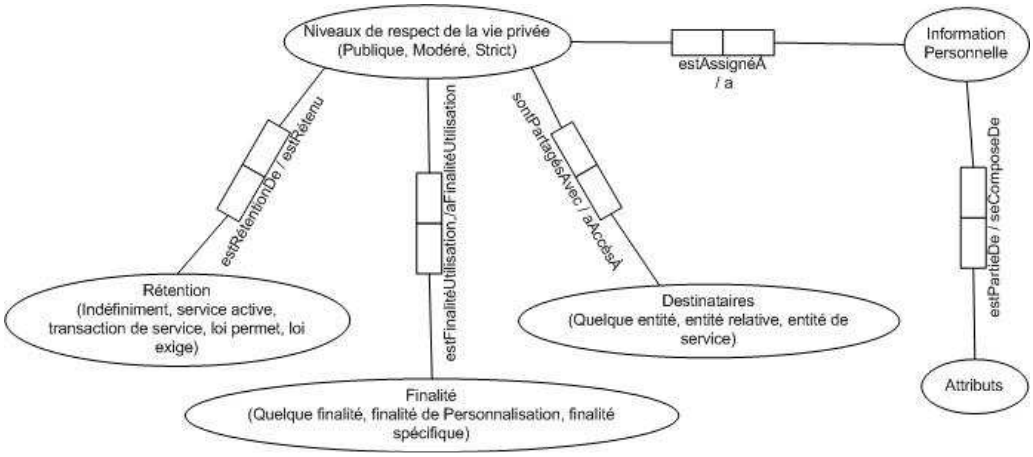


Figure 12. Ontologie des données personnelles associées à la vie privée

L'ontologie de la figure 12 indique que les informations personnelles se composent d'attributs et qu'un niveau de respect de la vie privée est associé à chaque catégorie d'information personnelle. De plus, chaque niveau de respect de la vie privée a des paramètres tels que le temps de conservation, la finalité d'utilisation et les destinataires.

Phase II : Dans cette phase, l'entité EgouvA envoie son ontologie d´information personnelle (elle joue le rôle de FI) aux entités FS (EgouvB). Les FS envoient en échange l'ontologie qui décrit les services à procurer aux utilisateurs. Le module du modèle de respect de la vie privée impliqué dans cette phase est : Spécifications de Méta-données (voir la figure 9).

Supposons qu'EgouvB est une entité de gouvernement fournissant les services de sécurité sociale avec les données personnelles suivantes pour chaque service :

Catégorie de service	***Service***	***Attribut demandé***
Assurance-vie	*Plan de retraite*	*Nom et prénoms, Date de naissance, État de santé*
Assurance-vie	*Plan d´éducation*	*Nom et prénoms, Titres professionnels*
Assurance Personnelle	*Assurance-accidents*	*Nom et prénoms, Nationalité*

EgouvA envoie l'ontologie d´information personnelle avec les conditions de vie privée à EgouvB et elle envoie en échange son ontologie de service avec l'information montrée dans le tableau ci-dessus.

Phase III : Une fois les ontologies créées et échangées dans la phase I et la phase II respectivement, le FI peut construire l'ontologie de politique de respect de la vie privée. Cette ontologie combine les services auxquels les utilisateurs sont autorisés à accéder avec les informations personnelles et les niveaux associés de respect de la vie privée. Le module du modèle de respect de la vie privée impliqué dans cette phase est : Spécifications de Méta-données (voir la figure 9).

L'ontologie de politique en matière de respect de la vie privée est présentée à la figure 13 dans le format ORM :

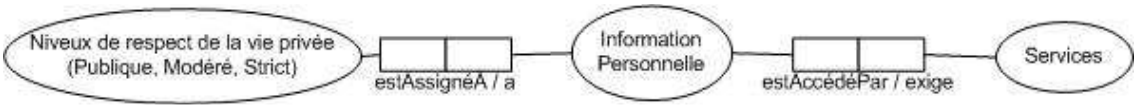


Figure 13. Ontologie de politique de respect de la vie privée

Cette ontologie est créée du côté d'EntA et elle indique que des attributs sont accédés par des services et que les informations personnelles ont associé des paramètres de respect de la vie privée. L´ontologie de politique est la base pour établir la politique en matière de respect de la vie privée finale.

Phase IV : L'ontologie de politique en matière de respect de la vie privée établie dans la phase III représente les spécifications par « défaut » associées à chaque attribut et aux services correspondants. Les utilisateurs peuvent à ce moment là exprimer leurs préférences en matière de la vie privée et modifier pour chaque attribut du niveau bas au niveau haut. Le module du modèle de respect de la vie privée impliqué dans cette phase est : Accès d´informations personnelle et des politiques (voir la figure 9).

La politique issue initialement de l'ontologie peut être modifiée par les préférences de l'utilisateur comme le donne en exemple la dernière colonne du tableau ci-dessous.

Catégorie d´information personnelle	***Attributs***	***Service permis***	***Niveau de respect de la vie privée***	***Préférences de l'utilisateur***
Données d´identification	*Nom et prénoms, Date de naissance, Nationalité*	*Assurance-vie, Assurance Personnelle*	*Public*	*Modéré*
Données académiques	*Titres professionnels, ID Professionnel*	*Assurance-vie, Assurance Personnelle*	*Modéré*	*Modéré*
Données de santé	*Maladies, Allergies, État de santé, Traitements médicaux*	*Assurance-vie*	*Strict*	*Strict*

Comme le niveau de respect de la vie privée pour la catégorie de données d´identification a changé de public à modéré, les attributs composant la catégorie peuvent être accédés seulement par des services connexes (les mêmes pour la catégorie de données académiques) ; pour sa part, la catégorie de données de santé est accédée seulement par le service qui demande l'attribut, dans ce cas-ci le service d'assurance-vie.

Phase V : Une fois que l'utilisateur a exprimé ses préférences en matière de vie privée, l'ontologie de politique est convertie en politique en matière de respect de la vie privée exprimée en un langage standard de contrôle d'accès. La politique de respect de la vie privée résultant est stockée et indexée avec l'attribut correspondant. Le module du modèle de respect de la vie privée impliqué dans cette phase est : Gestion de politique de respect de la vie privée (voir la figure 9).

Finalement, l´ontologie de la politique est traduite à une politique de respect de la vie privée. Une politique est un ensemble de règles qui sont évaluées en accord avec les paramètres de demande. La figure 14 montre les composants d'une politique ; l'objectif contient la ressource (quel attribut est demandé), et le sujet (quel service demande l'attribut). Les règles à évaluer utilisent d'autres paramètres de respect de la vie privée tels que la finalité et le temps de conservation des attributs.

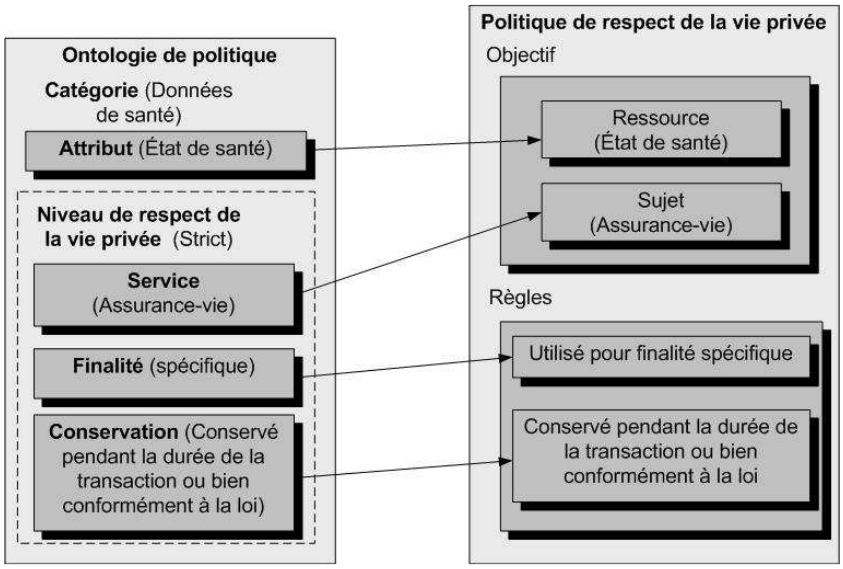


Figure 14. Traduction de l´ontologie de politique en politique de respect de la vie privée

Table of contents

Chapter 1

Introduction

Day after day, the number of online transactions is getting larger and larger. This has an impact on the way people interact among them and with diverse organizations. Most of our personal financial transactions are carried out by means of electronic banking; the number and type of items we can buy electronically are bigger and bigger, and diversified; distant learning is becoming an important means for providing formal knowledge; government entities are providing every day more citizen´s online services; and thus we can mention numerous examples.

The requirements for demanding online services are now more complex in terms of speed, availability, mobility and security. The entities providing services (known as Service Provider, SP) need to deploy new services as fast as the technological changes dictate, and those services must have excellent performance. The services must be available at any time and need to be accessed from anywhere; therefore, the access from mobile devices must be allowed. Finally, security aspects of the services are really important for guaranteeing their success. The SP must ensure the integrity and confidentiality of the information exchanged within the online transaction. Users need to be sure that their transactions are accurate and that their personal information is not compromised. That is why; privacy is an important factor to consider when deploying online services.

In order to guarantee the services´ requirements above exposed, every SP must properly manage and assign a digital identity to each user for accessing the services (digital identity is explained with more details in chapter two). A digital identity relates the physical entity with services that are allowed to access. When users access a great amount of online services – as nowadays it happens-, there is one digital identity assigned for each service; therefore, the number of digital identities are unmanaged causing a frustrating experience for the user. An additional drawback for users is that they have little or no control over their personal information regarding how it is handled in terms of privacy; that is, how it is collected, managed and exposed.

From the point of view of SPs, they need an Identity Management System (IMS) for managing the lifecycle of the digital identities of the users (Pope & Josang, 2005). The main problem for SPs occurs when they need to collaborate among them for providing coordinated services and multiple identities must be linked; for example, if the mobile TV service is provided by two SPs: the TV company and the mobile company, both IMSs must be integrated for linking the two independent digital identities so that the user can be treated as unique. Another possible interaction between both SPs is the exchange of personal information for service customization purposes. One possibility to overcome this problem is the creation of gateways so that IMSs can interoperate; it is a suitable solution when the number of IMSs is

small, but it is not scalable for a large number of them. A solid proposal for managing digital identities in a distributed environment is the deployment of a Federated Identity Architecture (FIA).

A FIA is a distributed IMS that allows different organizations to compose a Circle of Trust (CoT) for managing digital identities and sharing digital information (attributes of the digital identity) among their members (Dongwan & Gail-Joon, 2004). A CoT is built by means of business agreements and a common technological platform. FIA specification is supported by open technological standards that facilitate the interoperability of the commercial and open source solutions. Since almost ten years, there have been several standard initiatives for implementing them (these initiatives are explained in detail in chapter two). Nowadays, many manufactures are integrating federated extensions to their proprietary IMS solutions in order to be integrated into a federated environment. The entities composing a CoT can play the role of Identity Provider (IdP) or Service Provider; the IdP is responsible for authenticating the users and storing their identity data; meanwhile, the SP provides services to the users.

The FIA allows implementing identity services such as identity federation, Single Sign On (SSO) and attribute exchange. Identity federation is the link between two identities through the use of a common identifier known as pseudonym. SSO is the mechanism through which a user authenticates once with one IdP within the CoT and can access services from any other entity with no extra authentication. Attribute exchange allows interchanging attributes from different identities within the CoT.

As mentioned, FIA simplifies the intra and inter CoT attribute exchange; when a SP needs additional attributes from users in order to complete the service, it can request from the corresponding IdP of the same CoT (intra-CoT) or even from another CoT (inter-CoT).

The easy attribute exchange could represent a privacy issue if the corresponding measures are not taken. Regarding this point, the FIA has implemented some mechanisms that enhance in some extent the privacy (Varney, 2003). For example, all the communications between entities composing the CoT are encrypted and digitally signed preventing that any unauthorized entity could access the personal information exchanged. An additional feature that improves privacy within a FIA is that identities are federated (or linked) through the use of a pseudonym; a pseudonym is a digital identity that relates two digital identities, so that an IdP or SP does not know the name of the other identity to whom it is federated.

The measures above explained guaranteed a certain level of privacy in a FIA; however, privacy is not just a matter of technology. Privacy is recognized as a fundamental right for many years now, but the most important recognition dated from 1948 in the Universal Declaration of Human Rights (Privacy, 2006). The accelerated development of information technology has demanded the creation of privacy legal frameworks focusing on the protection of personal information with electronic representation. Since then, it has been developed basically two privacy legal models; the first one is a general model centered on the data themselves and the second one is a sector model centered on the person and the context where personal data are handled.

The general model comes from the European regulation represented by two of the most important international initiatives; the Guidelines on the Protection of Privacy and Transborder Flows of Personal Data emitted in 1980 by the Organization for Economic and Co-operation and Development (OECD, 1980) and the convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data created in 1981 by the Council of Europe (Europe, 1981). Among the main characteristics of the general model can be mentioned: public and private sectors scope, universal data protection, personal data controlled by users, prohibition of cross-border data flow, and the existence of national supervisory entity. Both initiatives propose fundamental privacy principles such as: *lawfulness* where personal data must be collected by legal means, *notification* which means that collection and use purpose of personal data must be notified, *quality* implying that personal data treatment must be accurate, adequate and not excessive, *consent* where user must approve how his personal data are handled, and *security* which enforces implementing mechanisms for guaranteeing the integrity, confidentiality and availability of personal data.

The sector legal model was mainly developed in the United States and it is focused on sectors with market auto-regulations; the government only takes part as sectorial supervisor. Examples of sectorial legal initiatives are the Health Insurance Portability and Accountability Act (HIPAA) emitted in 1996 by the U.S. Department of Health and Human Services (Health, 1996) and the Protection of Personal Data of Consumers dictated in 1974 by the U.S. Federal Trade Commission (Federal, 1974).

Mexico as well as other countries has developed privacy legal frameworks based on any of the two models or they have tried to adopt a hybrid model taking advantages of both. In the case of Mexico, it enacted the Guidelines for Personal Data Protection in 2005 (IFAI, 2005) and two constitutional amendments in 2008 (Senate, 2008).

The main problem within the context just described, is how the privacy principles expressed in any legal framework can be supported and enforced in a technological infrastructure such as the federated identity architecture. The present work proposes a privacy layered model for enforcing the privacy principles with complementary functionalities added to the basic operation of the IdP component in charge of the attribute sharing within the FIA. The basic idea is to protect personal information with privacy policies that are enforced each time an attribute is requested. The policies are created with internal and external legal requirements and privacy preferences of the user. Once the policies are created, they are associated to the personal data and evaluated each time an attribute is accessed.

The model is composed of three layers with components in each layer with specific functionality. The lowest layer deals with data specification (or metadata) converting the privacy requirements specified in natural language to formal expressions through the use of ontologies (an ontology is a formal specification that allows to represent concepts semantically); there are three ontologies needed to build the policies: the *personal information ontology* for representing personal data in a standard format independent of their heterogeneous source, the *privacy ontology* which expresses the privacy requirements and the *service ontology* which is received from the different entities to whom it will exchange attributes. All the ontologies are combined for expressing what type of personal data can be accessed from which services and with what privacy parameters; the resulted ontology composes the base for constructing the privacy policy.

The second layer supports building the privacy policy that is associated to the corresponding personal data, this policy can be accessed by users for consulting purposes and for expressing their privacy preferences. When the privacy policy is finished, it is ready to be evaluated by a module also from the second layer known as Privacy Policy Enforcement module or PPE. The PPE module is very important for the properly protection of personal data; it receives the attribute request and analyzes *what* attribute is requested, *who* is requesting the attribute (which service) and what is the privacy purpose of this request in terms of intended use and retention time. With the request parameters the PPE evaluates the policy and returns an answer allowing or denying the access.

The upper layer deals directly with the attribute release; it receives the attribute request from any entity of the CoT and redirects the request to the PPE for evaluating the purpose. If the response is affirmative then the attribute is retrieved from the storage and it is released to the requestor. This layer also provides an interface that allows users to access their personal data for validation and updating purposes. Through this interface, it is also possible to express the user´s privacy preferences and to verify the log transaction so that users and auditors know at any time how the personal data are handled in terms of privacy.

All the events generated by the modules of the model are logged: when the ontologies are built, each time a policy is generated or modified, every time an attribute is requested or modified. The log event is registered by a module that interacts with all the modules of the model.

The privacy model proposed in this work allows integrating privacy legal frameworks with an identity technological architecture in order to enforce the privacy principles dictated by such legal frameworks. To accomplish this work the model provides a set of functionalities such as:

- To convert privacy requirements expressed in natural language to formal ontological representation.

- To express heterogeneous personal data in a standard syntax.
- To allow users access to their personal data for validation and updating purposes.
- To create privacy policies with the participation of regulatory entities, organizations and users.
- To enforce privacy policies when attributes are requested.
- To audit all the events of attributes handling.

The architecture is modular, so that the modules can be integrated with the components of the IdP or they can be distributed and integrated with functionalities of other entities within the CoT.

The development of the project has consisted of a bibliography analysis related with the different identity management systems, specifically the federated identity administration system with their different initiatives proposed. The theme of privacy was covered from different points of view; from the point of view legal with the analysis of the main international regulatory frameworks and from the technological point of view analyzing the technologies that put in risk the privacy as well as those technologies that enhance it, mainly within the federated architecture. After that a correlation was made of the inherent privacy mechanisms of the federated architecture with the main privacy principles of the most important privacy legal frameworks. In this manner, the missing privacy aspects of the federated architecture that must be fulfilled by the model were defined. With those elements defined, the components and functionality of each module of the model were specified through the use of relation and events diagrams. Finally, the validation of implementation feasibility of the model was made with the development of a case scenario for the Mexican e-Government project of deployment of online services for the citizens.

The present document is structured as follows:

Chapter one (Introduction). - This chapter introduces the context, the problem and the solution proposed.

Chapter two (Digital Identity and Identity Management Systems).- It provides an overview of the concept of digital identity and the elements that compose it. The definition of Identity Management System is presented and the different models are exposed emphasizing their functionalities, advantages and disadvantages. The Federated Identity Architecture is presented as the most important solution for managing identities in a distributed and collaborative environment. The architecture and functionality of the main initiatives of FIA are explained in detail comparing their features and highlighting their challenges.

Chapter three (Privacy overview).- This chapter covers the theme of privacy from the technological and legal points of view. It describes the main legal frameworks regarding privacy developed in Europe and United States making emphasis on the different approaches taken by the two models. Similarly, the legal initiatives emitted in Mexico are presented and compared with the previous legal frameworks in terms of origin, focus and scope. From the technological point of view, some current technologies are discussed as they represent a risk to privacy or they enhance and preserve the privacy. The technologies representing a risk to privacy are known as Privacy Intrusion Technologies (PIT); meanwhile, those that enhance privacy are known as Privacy Enhancing Technologies (PET). In particular PET technologies related with policy languages are explained in details as they support defining, building and enforcing privacy policies.

Chapter four (Privacy Model for Federated Identity Management Systems).- In this chapter the favorable and unfavorable privacy aspects of the FIA are analyzed and mapped to show how they support certain privacy principles. The main contribution of this work is presented as a layered privacy model; it is proposed as a suitable solution for integrating a legal privacy framework into a FIA infrastructure. A detailed functionality of each module is presented indicating the components and interaction among them.

Chapter five (Case scenario: Privacy model for Mexican e-Government services).- This chapter presents the general context for e-government services mainly for the Mexican development in this area. Some activities of the Mexican Federal Public Administration are

exposed regarding citizen´s personal information handling. The implementation of FIA for e-government is explained as a suitable solution for managing citizen´s digital identities. Finally, a case scenario is presented where the privacy model proposed in chapter four can be applied for complying with a legal framework in an e-government service deployment.

Chapter six (Conclusions).- This chapter exposes the main aspects of this work emphasizing the challenges and the main contributions achieved during its development.

Chapter 2

Digital Identity and Identity Management Systems

The Internet has brought a huge increase in the number of on-line transactions among individuals and enterprises, accelerating the business relationships like B2B (Business to Business), B2C (Business to Client) and B2E (Business to Employee). At the same time, the requirements of the users have become more complex since they demand faster and more secure accesses, additionally with mobility facilities. Similarly, the technological convergence allowed multiple services and Service Providers (SP) to be integrated in order to offer joint services. For each accessed service, a digital identity must be assigned to the user by the SP, who must have an identity management system to handle the identity lifecycle such as creation, use and elimination (Pope & Josang, 2005).

Under this context, users feel uncomfortable handling several digital identities, one for each service, along with personal attributes. Besides, in most cases, users do not have control on the exhibition of their personal information, which constitutes a privacy problem that in some countries has legal repercussions. From the point of view of the SP, the identity management process represents a very high administrative load in financial and operative terms. Nevertheless, the main challenge that faces the SP is the difficulty to integrate with other SPs in order to be able to offer combined services and to handle a unique identity of the user. To deal with this problem, several Federated Identity Architectures (FIA) initiatives have appeared recently that propose a model of global identity management that allows to unify, to share or to link the digital identities of the users among different domains. After introducing FIA basic elements, three main FIA initiatives are described. For each initiative, its architecture, main components and operations are briefly explained. Finally, a comparison is made in terms of functionalities, and remaining issues and challenges are discussed.

2.1 Digital Identity

Identity could be defined as "the set of individual characteristics by which a thing or person is recognized or known" (Wordnet, 2008). If we transfer such definition to the digital world, we can define *Digital Identity* as the set of digital data that represents in a unique form an entity within application domain. In this context, an entity could be a person, a group of persons, an organization, a process or even a device, that is, any subject able to make a transaction. The elements that compound a digital identity are known as *Attributes*. Those attributes could be assigned, intrinsic to the entity or derived.

(Duran, 2003), proposes a layered structure for the attributes of a digital identity. Layer 1 contains the attributes associated with the entity and that are unconditional such as: name, birthday, eyes color, fingerprint, preferences, etc. Layer 2 relates with attributes assigned by other entities and they depend totally on the relationship with those entities, for example: driver´s license, employee number, passport number, among others. Layer 3 contains the attributes that are derived from the attributes classified into the others layers, examples or derived attributes are: *Residence District* derived from zip code or *Adult* derived from the birthday. The previous layered classification is important from the point of view of privacy, because it allows assigning different levels of privacy to each attribute.

The attribute or attributes that allow distinguishing a distinct entity within the context of a specific namespace are named *Identifiers*. Generally, identifiers are used to carry out the process of *Identification*, which consists in associating the digital identity with a real entity (Jean, 2004).

Authentication is the process of validating an identifier. For example, in a username/password system; the entity is identified by the identifier *username* and authenticated or validated through the *password*. Those elements used to carry out the process of authentication are called Credentials. A credential can be a password or the answer to a challenge (what he knows), or it may be constructed based on a smart card or a digital certificate (what he has), or any characteristics of the entity like his fingerprint, his eyes or his voice (what he is). The type of credential used during the authentication process depends on the business security requirements.

2.1.1 Anonymity and pseudonymity

The type of service provided determines the level of identified identity required (Windley, 2005). For many cases, it is needed a detailed identity information from the entity accessing the service, depending on the assurance level required for the transaction. However, there are services that do not require knowing the real identity of the entity demanding the service. In such case, the access is through an anonymous identity, which is a generic attribute that is not linked with a real entity.

An example of anonymous access is when a student wants to know the requirements for a scholarship within an educational service; the student enters the anonymous identifier "undergraduate" and then receives the information. When the identifier allows the service to distinguish among different entities and keep track of several sessions without linking to a real entity, the identifier is known as pseudonym. Identifiers used to access the Hotmail or Yahoo mail services represent examples of pseudonymity.

2.1.2 Digital Identity elements relationship

Figure 2.1 shows the existing relationship between the elements that compose a digital identity.

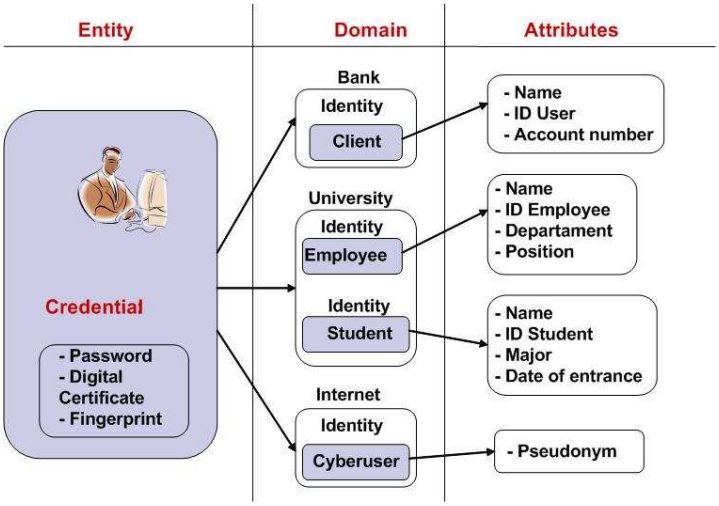


Figure 2.1 Relationship between the elements of a digital identity

In this scenario, the entity has many digital identities needed to interact with different application domains. An application domain is defined as the scope where the digital identity is valid, for example: a company, a hospital, a club, a university or the Internet. Note that an entity may have several identities within the same application domain. For instance, a professor could have both professor and student identities in case he takes continuous education classes.

Each identity is compounded by one or more attributes, additionally, credentials are associated to digital identities in order to validate the identifiers and carry out the process of authentication.

2.2 Identity Management System

Identity Management System (IMS) refers to the processes that handle the lifecycle of a digital identity, that is, the creation, handling and termination of a digital identity within an application domain. IMS also has to deal with the process of authentication, as well as the definition of the access control policy that an organization must fulfill in order to give access to protected resources (Subenthiran, Sandrasegaran, & Shalak, 2004).

2.2.1 IMS components and functionalities

An Identity Management System can be seen as a modular system composed of multiple services and components (Pato, 2003). Figure 2.2 shows the components of an IMS where the Digital Identity and Policy Repository is the core component linked to the operational components and lifecycle components:

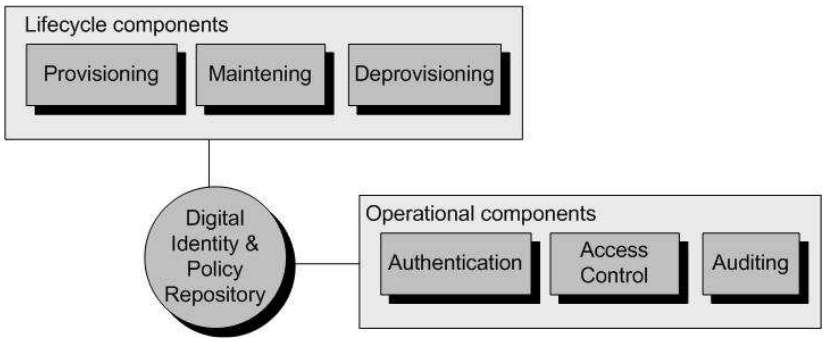


Figure 2.2 IMS components and functionalities

Digital Identity and Policy Repository.- It is the component where identity data, data model and authorization policies are stored.

Lifecycle components.- They are the components responsible for the lifecycle of digital identities from their creation, their maintenance to their elimination.

- Provisioning.- It is the functionality that creates the digital identities records and populates them with the appropriate attributes. Provisioning can be done through the action of an administrator, or it can be self-service which works well where there is little need to verify credentials.
- Maintaining.- Digital identities need maintenance due to the changes of the attributes and the lost or forgotten credentials by the users.
- Deprovisioning.- It is the action of removing identities from the system once they are at the end of their lifecycle. This process is very important from the point of view of security due to the risk that represents the unsuitable use of it.

Operational components.- These components manage the security issues of the digital identity, that is, they carry out the authentication process, the access control of protected resources and the auditing process.

- Authentication.- This process is the responsible for performing primary authentication of an entity. This component creates an authentication token which proves to other entities that the authentication has been performed. The authentication mechanism can use different techniques such as passwords, smartcards or biometrics depending on the security application context.
- Access control.- This module uses identity based information to control the access to protected resources. Authorization policies determine how information is manipulated.
- Auditing.- This component provides the mechanism to track how identity information in the repository is created, modified and used. This functionality is essential for analysis of regulatory compliance.

2.2.2 Identity Management System Models

Historically, IMS evolved from islands of identities, where each business unit of the organization managed its own identities with no integration (Isolated IMS). Later on, centralized solutions for unique handling of the users identities were implemented. Today, a number of ready-to-use products are available for organizations to implement their own private centralized solution (Centralized IMS).

Nowadays, the growing of on-line services provided in Internet and the increment of managed services (outsourcing) and business agreements between organizations, has resulted in an unmanaged number of digital identities that users must handle. That is why, new identity management models (User-centric IMS and Federated IMS) have appeared in order to facilitate the management of multiple digital identities (Pato, 2003).

The following sections describe the main IMS models in terms of their components and functionality, emphasizing their advantages and disadvantages.

For better understanding of the models, a definition of the following components is given:

- User.- Any entity that could be represented by a digital identity and that it is capable of doing a transaction, such as: a person, a group of persons, an organization or a process.
- Service Provider (SP).- An entity that provides a services to a user. Generally, it could be a Web site or a Web Service.
- Identity Provider (IdP).- The entity who manages the digital identity of the user and performs the authentication process.

Isolated Identity Management

In this model, each SP is responsible for managing the identity of each user. In this approach the SP deploys its own IMS with a complexity that depends on the functionality and trust that it wants to implement. Figure 2.3 shows the direct interaction between users and SPs for authentication process and service providing.

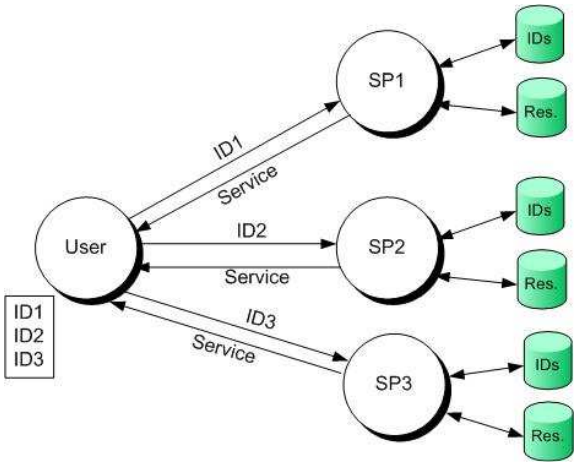


Figure 2.3 Isolated IMS

For SPs this model provides them full control of the identities, however, it represents an administrative cost and management burden. Another disadvantage from the point of view of SP is the difficulty to integrate its IMS with external IMS when it is necessary to deliver coordinated services among two or more organizations. For users, this model is appropriate when there are few services to be accessed, but it is really problematic when the number of SPs and consequently the number of digital identities increases.

The trust level is defined and implemented easily by the SP acting at the same time as IdP. The assurance level for the registration and authentication process is defined by the SP according to their risks assessment and sensitivity of the offered services. In this IMS model, the SP is in possession of the user´s information, therefore, he or she has not control over the information privacy.

The isolated IMS model has been used in Internet for many years as well as in organizations where the services provided are not coordinated.

Centralized Identity Management

In the centralized model, there is a unique repository of digital identities for management purposes. In this IMS model, the user may authenticate to all SPs using the same ID. Also, there is a possibility to implement a Single Sign On where the user authenticates once with the Identity Provider (IdP) and then can access all others SPs with no need for extra authentication. Figure 2.4 shows how the user can use the same ID for all SPs or only with the IdP if SSO is implemented.

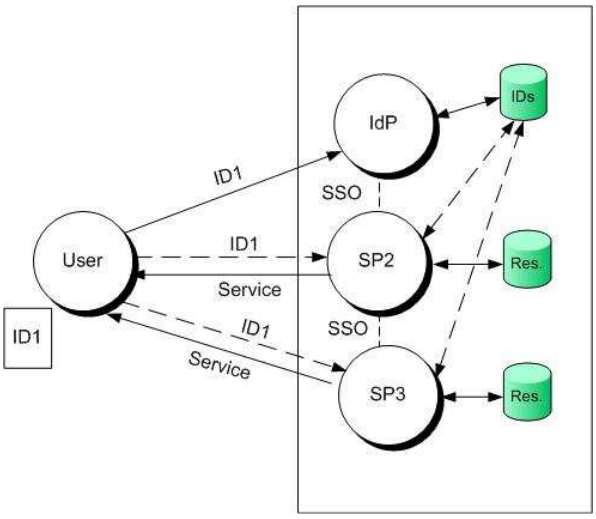


Figure 2.4 Centralized IMS model

This model provides an excellent user experience in relation to the facility to access multiple services with a unique ID and authentication process. For the SPs the task is simplified because all the identity management is carried out by the IdP which has total control over the user´s information. However, it also has some disadvantages like; the IdP and the IDs repository represent a single point of failure, the model is not scalable when the number of digital identities increases exponentially. Additionally, when there are many SPs is difficult to reach a high level of trust on a single IdP as the Microsoft Passport project showed when it tried to become the unique IdP of Internet.

The Centralized IMS model is used nowadays within organizations in order to simplify and reduce identity management costs. There are already many vendors which offer consolidated and robust identity management platforms.

Federated Identity Management

In the federated IMS model, digital identities of a single user that may exist across different SPs are linked or mapped by the use of pseudo identities. The SPs forming the federation (known also as Circle of Trust – CoT), establish a trust relationship among each other through a set of business agreement and technological framework. Figure 2.5 shows the elements of the federated IMS model and their relationships among them.

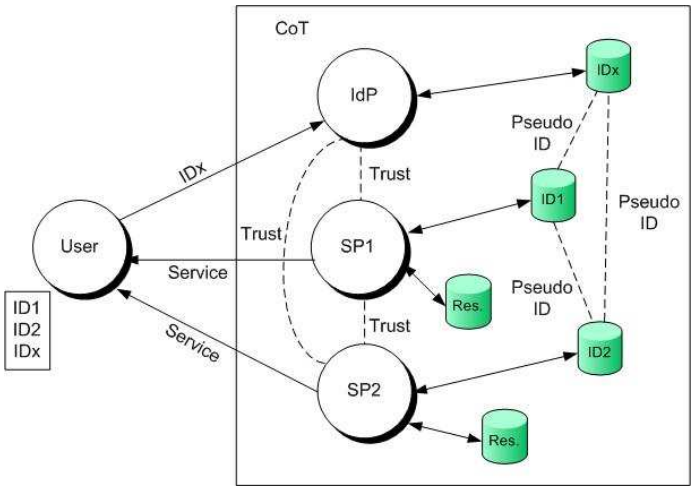


Figure 2.5 Federated IMS

Once the digital identities of the user are federated, he or she can authenticate with a specific service provider known as identity provider (IdP) and then have access to the rest of the SPs within the CoT without the need to authenticate again. Such functionality of SSO represents a better user experience when he or she must interact with multiple service providers.

For SPs, the federated IMS model facilitates the deployment of coordinated services among two or more SPs as far as they can treat users with a unique global digital identity, additionally, it represents a simplification in the identity life circle management and therefore a reduction in the administrative costs.

Nevertheless, this IMS model limits the scalability due to the strict and sometimes complex trust agreement that must be established among the members of the CoT. This complexity is accentuated when such trust relationship must be asymmetric due to the different risk exposure requirements of each member within the CoT.

For users it is not easy to have control over their personal information, additionally, they loose advantages of this model when it is necessary to access services from SPs that do not belong to the CoT.

Nowadays, diverse initiatives exist that define architectures of federated identities, nevertheless the three more important initiatives are: Liberty Alliance, Shibboleth and WS-Federation. Such initiatives are explained with more details in section 2.5

User-centric Identity Management

User-centric IMS was proposed to give users more control on their personal information by allowing them to choose identity providers independently of service providers. Meanwhile in the federated IMS the IdP works in the interest of the SP, the goal of a user-centric approach is to enable the creation of IdPs who operate in the user´s interest. In this IMS model, the user can select any IdP based on its security and policies practices and use his identities with different SPs. The SPs don't need to establish a trust relationship between them in order to provide service to the end user. Figure 2.6 shows the elements forming this IMS model and the interactions among them.

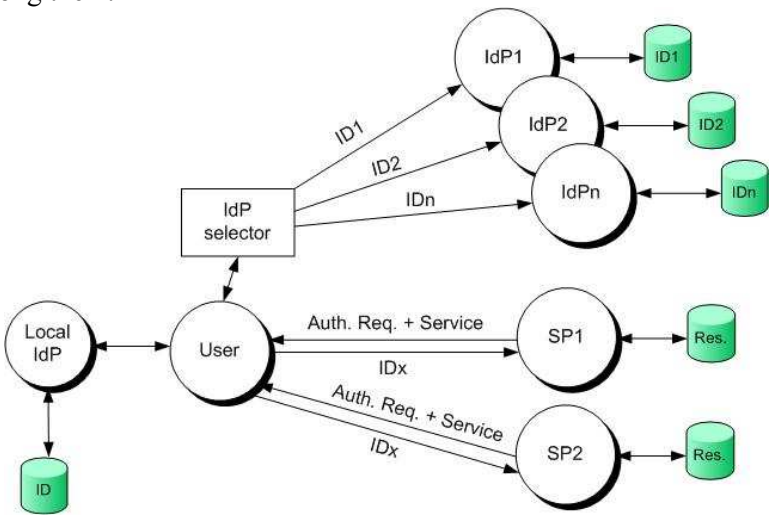


Figure 2.6 User-centric IMS

In this model, each SP specifies the authentication requirements to the user, who can select the appropriate IdP to obtain the ID information in order to get access to the SP service. The user-centric approach allows the user to implement his own IdP (personal IdP) which can be implemented in a PC or any mobile device, such functionality is appropriate when the user provides directly personal information to the SP. A disadvantage from the point of view of SP is the difficulty to interact with other SPs in order to provide coordinated services to the end user.

There are already some serious initiatives supporting this IMS model, such as the open source Eclipse Higgins Project and the CardSpace Project from Microsoft.

2.2.3 Identity Management Models comparison

Table 2.1 shows the IMS comparison in terms of four main characteristics; user experience which refers how easy the user can handle multiple identities, personal information control which defines the level of privacy controlled by the user, IMS integration which specifies how easy two or more IMS can interoperate in order to offer coordinated services, and scalability which determines how big can be the model in terms of number of users, service providers and identity providers.

IMS model	User experience	Personal info. control	IMS integration	Scalability
Isolated	Acceptable (with few identities)	Low (control by SP)	Very low (heterogeneous technologies)	High
Centralized	Very good (with SSO)	Low (control by IdP)	Moderated (with use of IMS gateways)	Low
Federated	Very good (with SSO)	Low (control by IdP)	High (common technological platform)	Medium
User-centric	Good (with adequate interfaces)	High (control by user)	Low (Based on user technology)	High

Table 2.1 IMS models comparison

The isolated model has traditionally dominated Internet, it is very scalable but it is extremely limited when the user has multiple identities, also it is not easy for organizations to integrate their IMS in order to provide coordinated services. Isolated IMS model is applicable only to small and controlled environments without external interaction.

The centralized IMS model was proposed as an alternative model to simplify the management of digital identities by the consolidation of such identities in a centralized identity repository. This model simplifies the lifecycle of digital identities, but it introduces some shortcomings as it constitutes a single point of failure and it does not scale well with an increasing number of identities. The centralized IMS model is targeted to organizations or to a limited group of organizations that relay identity management on a centralized entity.

The federated IMS was designed to facilitate the exchange and sharing of digital identities among organizations which have built a trust relationship, this architecture simplifies the identity management and enhances the user experience, however, it limits the scalability due to the restricted business agreements among the entities. In all these previous models, personal identity information is maintained by the service providers or identity providers; therefore, the user has null or minimum privacy control over its personal information.

The last IMS model, the user-centric model tries to alleviate the problem of privacy by delegating the control of personal information entirely to the end user. Despite the improvements of privacy issues, the user-centric model has the disadvantage of IMS integration between service providers in order to offer coordinated services.

Federated and user-centric IMS models contain the best characteristics to fulfill the actual global identity management requirements. The federated IMS is targeted to a scenario where the

type of transaction requires a high level of trust, such as financial or B2B environments, meanwhile user-centric model applies where the user needs to control personal information but a high level of trust is no required. Without doubt, these two models will interoperate in order to work in a majority of application scenarios.

Without concerning the IMS model used, there should be a trust relationship between the entities exchanging identity information. The following section will describe the main trust models and how they are related with the IMS models above described.

2.3 Trust Models

A fundamental characteristic in any IMS model is the level of trust relationship that must exist between the components exchanging identity information. In this context, a trust relationship can be defined as the combination of a business agreement and an authentication mechanism established between them. Business agreement and authentication mechanism can be achieved in an indirect or direct way, depending on an intermediary element is used or not respectively (Josang, 2005). When a direct business agreement is established, the trust model is known as Pairwise Trust Model, if the business agreement is set up through an intermediary entity, then the model is named Brokered Trust Model. However, for particular transactions, it is not necessary to establish a business agreement, the overall trust may be derived from the enrollment and participation in a shared authentication infrastructure (Linn, 2004).

The authentication mechanism in each trust model can be direct or indirect. Direct authentication trust is achieved by the direct usage of cryptographic keys between the two elements; meanwhile indirect authentication trust is established through one or more intermediaries in such a form that an authentication path can be derived for the two end entities.

The following subsections will describe the main characteristics of such trust models and their relations with the IMS models previously described.

2.3.1 Pairwise Trust Model

In this model, there is a direct business agreement among the entities exchanging identity information. From Figure 2.7, entity A has a list of all entities to whom it has a direct business agreement (entities B, C and D). In case that it has also a direct authentication mechanism with them, they are included in another list of authentication pairs (entity B and C). If the authentication mechanism is indirect (entity D), then an intermediate entity which is used to build the authentication trust path is included in such list (broker AB).

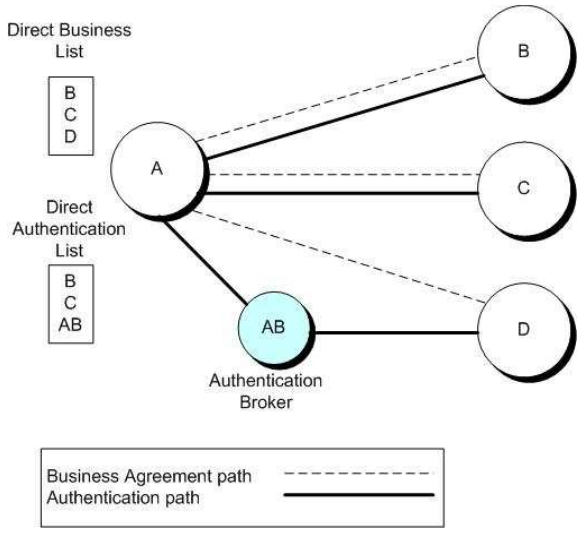


Figure 2.7 Pairwise Trust Model

In Figure 2.7, if entity A represents an SP (SPA) and entities B, C, and D represent IdPs (IdPB, IdPC and IdPD respectively), there are business agreements signed among the SP and all IdPs. The business agreements specify information such as; the type of information exchanged, the authentication mechanism used, set of privacy policies, etc. If IdPB or IdPC send security messages (authentication assertion, attribute response, federation request, etc) to IdPA, it has in its authentication list the keys required to authenticate such messages.

However, if there is a security message sent from IdPD to SPA, the SP needs to build an authentication path to obtain the key required to authenticate the message. The indirect authentication mechanism can be implemented with technologies such as PKI or Kerberos infrastructure, where one or more authentication brokers act as intermediaries between the two end entities.

2.3.2 Brokered Trust Model

In this model, there is no direct business agreement between the entities exchanging identity information. The brokered trust model is widely used when transactions span multiple administrative domains. Overall trust is built by the combination of all direct business trust through the entire path. Figure 2.8 shows how the business trust is built in this model.

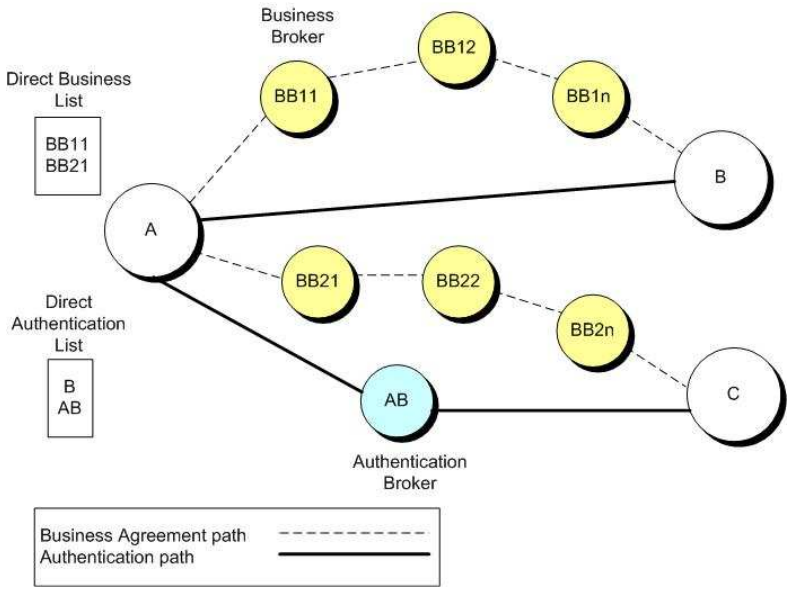


Figure 2.8 Brokered Trust Model

Similarly, if entity A is an SP (SPA) and entities B and C represent IdPs (IdPB and IdPC respectively), the business agreement among them is reached through intermediary entities (known in the diagram as business brokers -BBs), it must be possible to build an indirect business agreement path by the verification of each direct business agreement established between each pair of brokers. SPA contains in its business list a reference of the BBs with which it has direct business agreements (BB11 and BB21 in the example). The business brokers are located generally in different administrative domains.

This trust model accepts a direct or indirect authentication path as explained in the pairwise trust model.

2.4 Federated Identity Architecture

A Federated Identity Architecture (FIA) is a group of organizations that have built trust relationships among each other in order to exchange digital identity information in a safe way, preserving the integrity and confidentiality (privacy) of the user personal information (Dongwan & Gail-Joon, 2004).

The FIA allows the sharing of identity information between different domains, having guaranteed their integrity and confidentiality, such functionality will facilitate to the companies to interact and to be able to offer services of an easy and secure way. The federated identity management will promote the development of businesses between entities (B2B, B2C and B2E).

The FIA basically involves Identity Providers (IdP) and Service Providers (SP) in a structure of trust by means of secured communication channels and business agreements (Bhargav-Apantzel & al, 2005).

IdP manages the identity information of the user and does the authentication process in order to validate his identity. Within a FIA there could be one IdP (centralized model) or several IdPs (distributed model). The centralized model has the advantage that the identity information is not disseminated, facilitating its confidentiality and integrity, but it could represent a bottleneck and a single point of failure. In the distributed model, the authentication process can be done in any IdP, providing flexibility and load balancing.

However, this approach requires more complex and secure mechanisms to exchange, and to manage the identity information guarantying its integrity.

SP provides one or more services to the users within a federation. The enforced access control policy protects the services themselves by granting access to authorized users. This access control policy is established when the federation is formed.

The FIA must fulfill the following main functionalities from the point of view of users, identity providers and service providers:

- **Single Sign-On (SSO).-** SSO allows users to authenticate to an IdP and then to access services provided by several SPs with no extra authentication.
- **Attribute exchange.-** As mentioned before, for access control enforcement, SP needs authenticating an identity, but also getting additional attributes to provide personalized services. This means, that the FIA must facilitate attribute exchange between IdP and SP.
- **Personal information privacy**.- Confidentiality and integrity of the user´s personal information must be guaranteed in such a way that the exposure of the identity attributes must be controlled by the user.
- **Identity lifecycle management**.- Whether the model is centralized or distributed, the creation, maintenance and elimination of a digital identity must be simple and must not represent high operational costs.
- **Standardized architecture**.- The FIA must be based on standards for an easy integration of new SPs and IdPs.

2.4.1 Federated identities

In a federated architecture, each IdP and SP may have a digital identity for a specific user. Those identities must be linked in order to implement SSO and attribute sharing facilities. When digital identities are federated, a federated identifier is created for each pair of IdP and SP, linking two digital identities. The federated identifier can be dynamic, that is, for each user's session, or may be pre-established for a long period of time.

The federated identifier is basically a pseudonym that preserves the privacy of the real identity at each entity, SP or IdP. A business agreement between the entities, establishes the way the identities will be federated (structure of pseudonym, transient or permanent identifier) as well as if some attributes are exchanged at federation time.

When there is no digital account at the SP, a federated identifier could be used for anonymous access using a generic pseudonym or attribute, for example the generic pseudonym "visit01" could federate the identity ID1 at the IdP with the "Guest" account at the SP. However, the SP does not know the existence of the identity ID1. This type of federated identifier is transient and ends when the session finishes.

Generally, when two digital identities are federated, the identifier or pseudonym used is a permanent one, that is, it lasts for several sessions or until the identities are defederated. However, it is recommended to change the federated identifier periodically in order to preserve the identity privacy.

Figure 2.9 shows the federation of identities using a permanent pseudonym. The identity table at the IdP shows how the identity ID1 was associated to the federated identifier 65ER4589 when the identity ID1 was federated with the identity ID2 at the SP1. Similarly, the identity table at SP1 shows the relationship of the local identity ID2 with the federated identifier 65ER4589, which is used for all identity information transmitted between IdP and SP for that particular user.

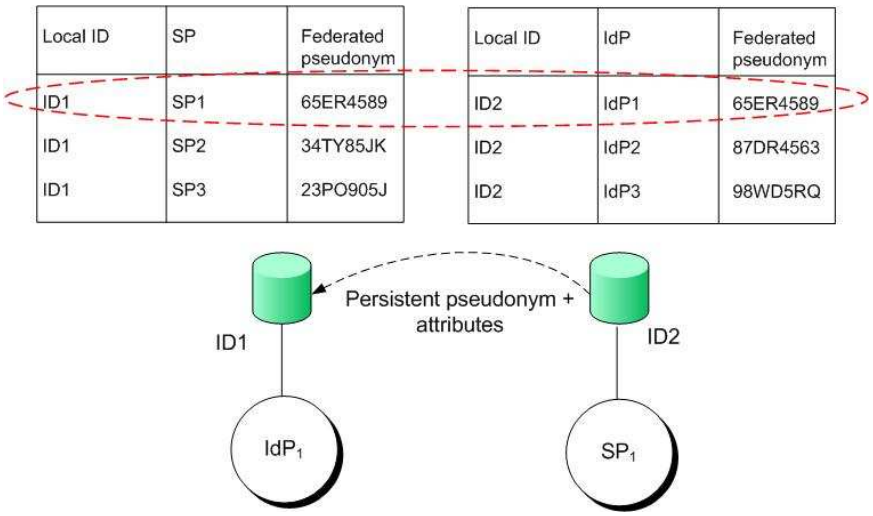


Figure 2.9 Identity federation using permanent pseudonym

The pseudonym has a local meaning. Therefore, the IdP and the SP only know the local account and the pseudonym, but they do not know anything about the remote digital identity. When two or more SP need to interact to provide a service to a common user, they use the identifier mapping protocol to query the IdP to obtain the federated identifier to be used at some other SP.

2.5 Federation Identity Protocols

The development work of OASIS regarding federated protocols was shown with the first standard named SAML V1.0 (Security Assertion Markup Language) in November 2002. SAML V1.1 followed in September 2003 and has seen significant success in many organizations of different segments.

In July 2002, Liberty Alliance (a consortium of many organizations and industries) released its first set of specifications named Identity Federated Framework (ID-FF 1.0) based on SAML V1.0. By January 2003, some improvements were made and the ID-FF 1.1 specification was released. This period of time was known as the phase I of the Liberty Alliance project.

In 2003, an initiative promoted by a group of Universities from the project of Internet II launched the federated protocol specifications named Shibboleth which main objective was to share academic resources preserving the privacy of the user identity. This specification was based on SAML V1.0.

In November 2003, the phase II was released and included the ID-FF 1.2 specification and the Identity Web Service Framework (ID-WSF 1.0), both specifications operating on SAML V1.1. In a similar way, Shibboleth V1.2 was released in April 2004 using SAML V.1.1 as a base framework.

Until this point, all federated protocols were non-interoperable and non-backwards compatible.

SAML V2.0 (released in March 2005), unified the building blocks of federated identity in SAML V1.1 with input from both higher education's Shibboleth initiative and the Liberty Alliance's Identity Federation Framework.

New releases of Liberty and Shibboleth adopted SAML V2.0. as the federated protocol framework to build all of their functionality specifications. As such, SAML V2.0 is a critical step towards full convergence for federated identity standards.

Figure 2.10 shows the identity protocols evolution and the dependencies of each version initiative:

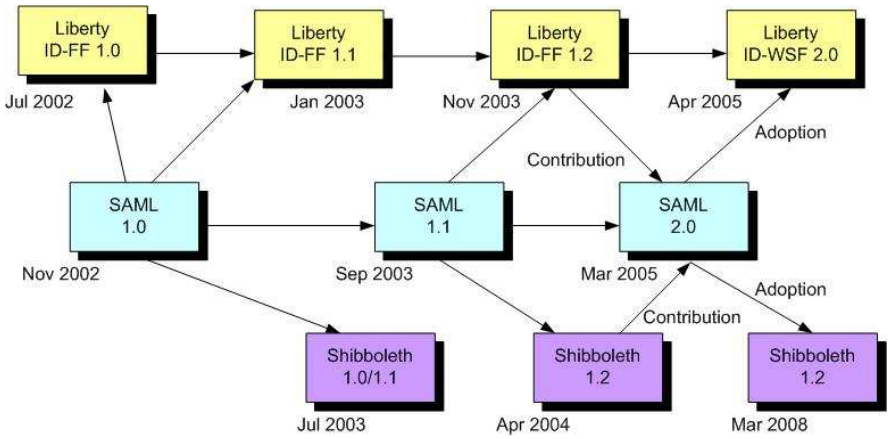


Figure 2.10 Federated Identity protocols evolution

Meanwhile the previous federated protocols have evolved; the WS-Federation specification developed by Microsoft was released at the end of 2005 as important component of the WS-Security framework for Web Services. Without doubt, SAML 2.0 and WS-Federation will be the most important standards widely used for deployment of federated architectures.

2.5.1 Security Assertion Markup Language (SAML)

SAML is a standard developed by the Security Services Technical Committee of the Organization for the Advancement of Structured Information Standards (OASIS), SAML defines an XML-based framework for communicating user authentication, entitlement, and attribute information between on-line business partners (Ragouzis, 2006).

SAML provides the fundamental services required by federation architecture; Web SSO, identity federation, attributes exchange and Single Logout (SLO). In order to accomplish such functionalities, SAML has a hierarchical architecture composed by: *Assertions*, *Protocols*, *Bindings* and *Profiles*.

- **Assertions** are affirmation made by a SAML authority regarding identity information. It could be an assertion of authentication, attribute or authorization.

- **Protocols** are request/response pairs in order to exchange assertion and manage identities.
- **Bindings** detail how SAML messages can be carried over common transport protocols.
- **Profiles** define how to combine assertions, protocols and bindings to support a particular application.

Figure 2.11 shows the hierarchical architecture of SAML.

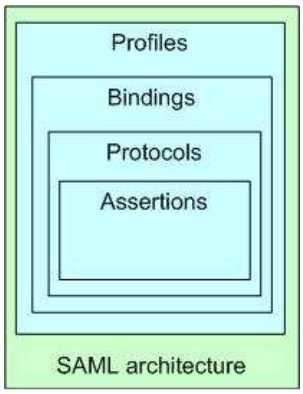


Figure 2.11 Hierarchical SAML architecture

Assertions

SAML assertions allow an entity (named SAML assertion party) to assert security information in the form of statements about a subject. The statements could be: Authentication statements which are used to authenticate the user and the specific time at which the authentication took place, Attribute statements that contain specific identifying attributes about the subject and Authorization decision statements which define something that the subject is entitled to do.

The structure of an assertion is built by four parts; the first block defines the SAML version, the time when the assertion was issued and the ID of the issuer; the second part identifies the subject which can be expressed in different formats (X.509, Windows domain, Kerberos, pseudonymous, among others); the third part specifies optional conditions under the assertion is valid, and the last part is the statement which corresponds to the identity information asserted. Figure 2.12 shows the general structure of an assertion and exemplifies an authorization assertion.

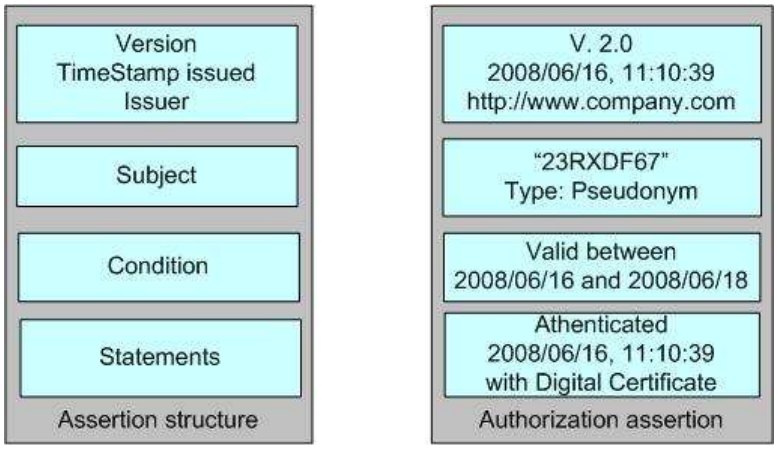


Figure 2.12 General assertion and Authorization assertion

Protocols

A SAML Assertion Party (SAML AP) is an entity which creates and sends assertions. Meanwhile a SAML Relying Party (SAML RP) is the entity which consumes them. In order to exchange such assertion, a set of request/response protocols have been already defined (Ragouzis, 2006):

- Protocol to request the SAML AP to authenticate a principal and to generate an authorization assertion with the information about the authorization context; when and how the principal was authenticated.
- Protocol to request the SAML AP an assertion containing attribute information regarding the principal.
- Protocol to terminate an active session associated with a principal.
- Protocols to modify or map the name identifier used to refer to a principal.

Bindings

Bindings specify how SAML messages are transported through a communication protocol. Basically, two mechanisms were selected; HTTP and SOAP. HTTP is the HyperText Transfer Protocol used for Web browsing. HTTP is used when communicating parties exchange SAML messages through an HTTP user agent. This protocol uses two different methods to transport SAML messages; the first one is the HTTP redirect, which uses URL directives to encode and sign SAML assertions within the header, the second option is the POST functionality of HTTP, where the SAML assertion could be encoded and transported as a HTTP form control. When the HTTP agent does not support directly the transmission of the SAML message, a reference to the message named Artifact could be sent via HTTP redirect or HTTP Post. After receiving the artifact, the communicating parties use a specific protocol to exchange directly the SAML message.

The second communication protocol used is SOAP (Simple Object Access Protocol), which is a peer-to-peer protocol for exchanging structured and typed information between peers in a distributed environment. It uses XML technologies to define an extensible messaging framework providing a message construct that can be exchanged over a variety of underlying protocols.

The decision of which bindings to use is typically driven by configuration settings at the IdP and SP systems.

Profiles

Profiles are usage cases where different assertions, protocols and bindings are combined in order to accomplish a specific application.

SAML specifications, define a set of basic profiles but additional profiles could be defined. Some of the most important profiles already defined in the latest SAML version are:

- **Web Browser SSO**.- Defines how messages and assertions are used to achieve SSO with standard web browsers.
- **Enhanced Client and Proxy (ECP).-** Defines a profile where specialized clients or gateway proxies can use the SOAP binding.
- **Single Logout Profile**.- Defines how to implement single logout functionality using different bindings.
- **Assertion Query/Request.-** Defines how to obtain SAML assertions over a binding, such as SOAP.

As an example, the Web Browser SSO profile is described in the next section, showing the messages interchanged among the communicating parties (User, IdP and SP).

Web Browser SSO Profile

By implementing this profile within a federated architecture, a web user may authenticate at the IdP once, and then access one or more SPs within the federation without authenticating again during the session lifetime. The process can be SP-Initiated or IdP-Initiated, depending where the process starts, at the SP or IdP respectively. As an example, for a SSO SP-Initiated the bindings given in Table 2.2 are possible.

SP -> IdP **Authentication Request**	**IdP -> SP** **Authentication Response**
HTTP Redirect	HTTP Post
HTTP Post	HTTP Artifact
HTTP Artifact	

Table 2.2 Bindings combination for a SSO SP-Initiated profile

In HTTP Redirect binding the SAML message is transmitted within the HTTP header as URL parameters. The HTTP Post binding uses a HTTP form control to send the SAML message. In the HTTP Artifact binding a small reference named Artifact is sent within the HTTP header. Such Artifact is used to request the SAML message.

The diagram depicted in Figure 2.13 shows the interactions and messages exchanged for a Web SSO SP-Initiated using HTTP Post binding for the authorization request and the HTTP Artifact binding for the authorization response.

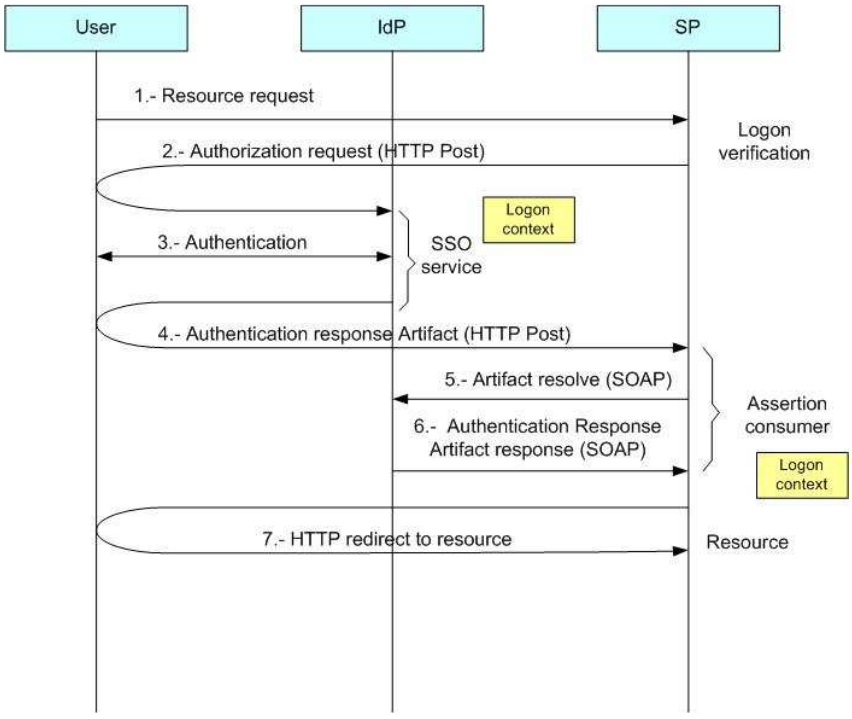


Figure 2.13 Web SSO profile using HTTP Post/HTTP Artifact bindings

1.- The user requests the access to any resource at SP, 2.- The SP verifies if the user has an active session, if not, then send an authorization request to the proper IdP in SAML format using the HTTP post binding, 3.- The SSO service at the IdP challenges the user for authentication as requested by the SP, if successful, the SSO service creates a logon context of the active session, 4.- The SSO service builds an authentication response, in this case it uses the Artifact method and sends to SP (assertion consumer service) a reference to the SAML

message, 5.- The SP uses the artifact reference to request the authorization response using the SOAP protocol, 6.- The IdP verifies the signature of the artifact resolve message and sends the authorization response to the SP for the assertion consumer, a logon context is created at the SP, 7.- Finally, the authorization response is evaluated for access control purpose and the web resource is granted to the user.

This profile is used when the SAML authorization response can not be transported directly into the user web agent and the authorization response message is send directly between the IdP and SP in a back security channel, which is an encrypted and authenticated communication channel between the IdP and SP.

2.5.2 Liberty Alliance

Liberty Alliance is a group of more than 200 companies from diverse sectors. It was launched in 2001 with the objective to establish a technological, business and policy framework for implementing a Federated Identity Architecture (Watson, 2003).

The Alliance developed a business guide to help companies converge towards a business agreement and conform to a federated architecture focusing on feasibility, risk, mutual trust and compliance aspects.

Liberty Alliance looks for working with other organisms of standardization that are working in similar initiatives to integrate their standards and to obtain the interoperability of their architectures.

Liberty Alliance does emphasis in aspects of the identity privacy and anonymous issues, specifying access control mechanisms on the identity attributes and handling pseudonyms to hide real identities in a federated context.

Architecture

Liberty Alliance is a framework that includes a set of technical and business specifications for establishing a Federated Identity Architecture. Its architecture shown in Figure 2.14 includes three modules that operate on technological open standards developed by organisms like: OASIS, W3C and IETF*.

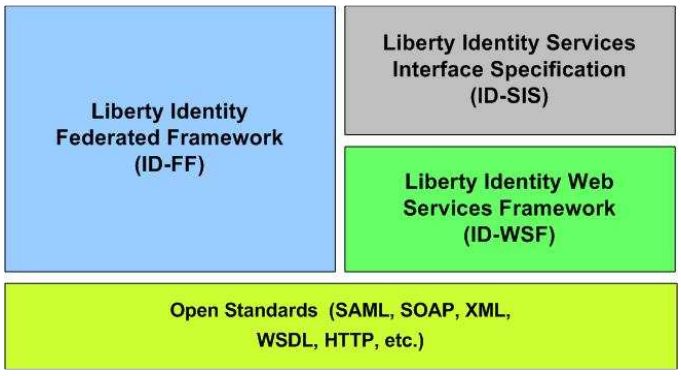


Figure 2.14 Architecture of Liberty Alliance

ID-FF (Identity - Federation Framework) is a set of specifications targeting identities federation and management. This module composes the fundamental part of the architecture, defining a set of functionalities like: account linking (identity federation), session management (Single Sign On and Single Log Out), affiliation (capacity to select the IdP for identity federation).

* OASIS (Organization for the Advancement of Structured Information Standards), W3C (World Wide Web Consortium), IETF (Internet Engineering Task Force)

ID-WSF (Identity - Web Services Framework) specifies a framework for Web Services in order to create, discover and request identity services. ID-WSF also operates on open protocol standards (Jonathan, 2004) and supports the following functions: attributes sharing (with possible previous authorization from the user), discovery of services, security mechanisms to transmit messages, etc.

ID-SIS (Identity - Services Interface Specification) serves to build security services of higher level (applicative services) based on the ID-WSF framework. Examples of ID-SIS services include: personal information request, geo-location services, directory services, etc.

Elements and operation

Liberty Alliance defines a Circle of Trust (CoT) to which SPs and IdPs adhere by signing a business agreement, in order to support secure transactions among CoT members. As depicted in Figure 2.15, each member of CoT might know a user under distinct identities. All identities are related or federated in such a way that the authentication process can be performed by any CoT member. In that sense, Liberty Alliance is said to be distributed because any SP implementing IdP within the CoT may authenticate a user.

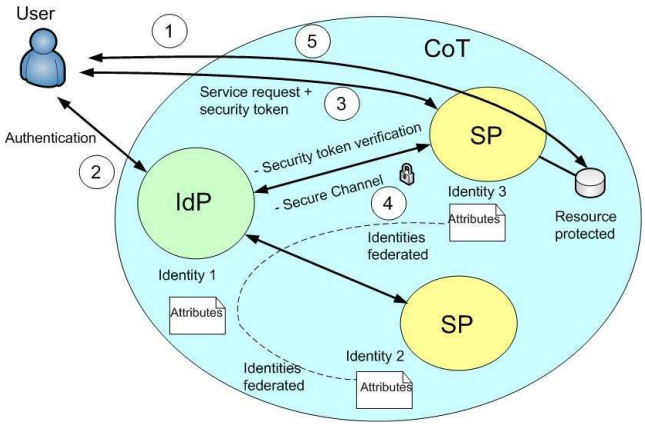


Figure 2.15 Circle of Trust (CoT)

For the user to access any service inside the CoT (1), the SP asks the user to select an IdP, and the user is redirected to this IdP for authentication (2). The IdP authenticates the user and assigns a "security token" with identity information which is next forwarded to the SP(3); the "security token" is verified between the SP and IdP in a back secured channel (4), and in case of validity, access to the service is granted (5). If the SP requires additional attributes, they are requested to the IdP through the secure channel.

The CoT model demands that SP trusts the IdP, thus, it requires a secure communication infrastructure that guarantees the integrity, confidentiality and non repudiation of the interchanged messages. The incorporated security mechanisms in the specification of Liberty Alliance include security in the communication channels as well as security in message exchanges. The secure communication can be implemented by means of current standard protocols such as: TLS, SSL and IPsec. These protocols implement authentication mechanisms between SP, IdP and users before initiating the message exchange (Watson, 2003).

Functional characteristics

The reference framework of Liberty Alliance (ID-FF, ID-WSF and ID-SIS) includes a set of functional characteristics, among which we can mention the following ones:

Circle of Trust (CoT) building. - The components of the model (SP and IdP) must establish a technological, policy and business agreement to be able to conform federated identity architecture.
Federation of identities. - It is carried out by means of the identities linking and sharing under the control and authorization of the user.
Privacy handling. - It defines access controls policies on the identity attributes and uses pseudonyms and encryption technologies.
Sessions Management. - Handling of decentralized Single Sign On and Single Sign Out by means of authentication requests and answers based on SAML assertions.
Secure network architecture. - It specifies a secure communication channel by coding and authentication protocols, in addition by means of the digital signature of interchanged messages.
Discovery and interchange of attributes. - Facility to discover security services related to a user and the interchange of identity attributes between service providers and identity providers.
Interaction with the user. - Specifications to interact with the end user through direct requests of attributes or permission to access its attributes handled by some identity provider.

2.5.3 Shibboleth

Shibboleth is an academic initiative of University members of Internet 2. Its objective is to facilitate the collaboration and access to protected resources among institutions without using external or temporary accounts. Some applications that could take advantage of this solution are: access to library database information, distance learning courses, collaborative applications for project development, etc (Scavo & Cantor, 2005).

In Shibboleth, information relative to the users digital identity is managed by the institution to which they belong. When a user requires access to the resources located in another institution, the identity attributes are sent along with the request but only attributes previously agreed to be shared might be communicated. These attributes are finally used to make decisions of accepting or rejecting user's access request according to the local access control policy. The main interest is to distinguish between users belonging to an institution and students from a specific course. Thus, it is not necessary to send the real identity of the user, and so privacy of personal information may be guaranteed in Shibboleth.

Architecture

The architecture is also built upon open standards such as: HTTP, XML, SOAP, and SAML (Cantor, 2005).

Figure **2.16** depicts the components composing the architecture and their supported services.

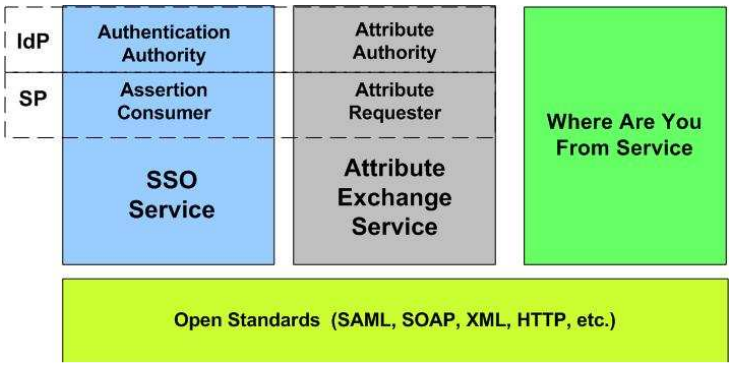


Figure 2.16 Shibboleth architecture

The components include:

Authentication Authority.- It emits authentication assertions to the Service Provider, this module does not specify how the user authentication must be done, but it works in coordination with the local authentication system. This component is part of the IdP entity.

Assertion Consumer.- It processes the assertions sent by the Authentication Authority of the IdP and finally generates a security context in the SP (secure communication channel) and redirects the client towards the resource protected by the SP. This component is part of the SP entity.

Both components conform the Single Sign On (SSO) service within the Shibboleth architecture.

Attribute Authority.- It emits attributes of the user to the SP under a mechanism of mutual authentication based on SSL/TLS and signed SAML messages. The attributes only will be able to be emitted by means of access policies defined by the IdP and the user, through this mechanism, the privacy of the personal information is guaranteed. This component is part of the IdP entity.

Attribute Requester.- Component that requests for attributes to the IdP by means of a secure messages interchange, it belongs to the SP entity.

Attribute Authority and attribute Requester provide jointly the attribute exchange service.

WAYF (Where Are You From).- WAYF is an optional component that enables the SP to locate the user's IdP subscription. WAYF is pretty like a directory that interacts with the user for the selection of the IdP that conducts the authentication operation.

Elements and operation

Shibboleth consists of three elements: Origin (Identity Provider), Target (Service Provider) and optionally the WAYF (Where Are You From). The Origin maintains users' accounts (credentials and attributes) and carries out the authentication function. In addition, it generates authentication and attribute assertions towards the Target. The Target manages the protected resources and controls its access based on the assertions of identity emitted by the Origin. The WAYF if implemented, allows the user to select the Origin in charge of the authentication process (Cantor, 2005). Figure 2.17 shows the relationship and operations between Shibboleth components.

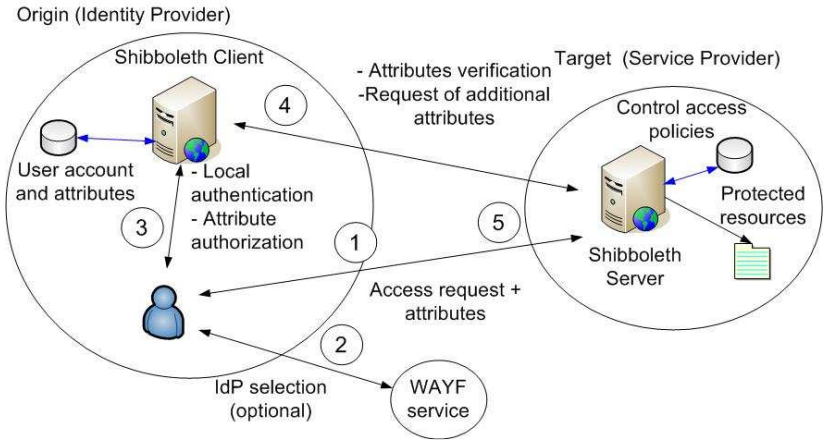


Figure 2.17 Elements of Shibboleth

When the user needs access to a protected resource located outside his organization (1), the Target asks the user to authenticate himself. Usually, the Origin or IdP is the organization to which the user belongs, optionally, the WAYF service can be used to select the Origin (2). When the user is authenticated (3), the Origin assigns attributes which are presented to the Target. These attributes are proved as authentic since they are delivered through a secure communication channel (4). In case of successful authorization by the access control policy, access to the resources is granted (5). In some cases, additional attributes might be required in order to provide the services, so the needed attributes are requested to the Origin. These attributes are sent only after getting the user's authorizations. Within the architecture of Shibboleth, the privacy of the personal information is very important.

As it can be seen, the identity information of the user resides solely in the Origin, but some attributes might be communicated to the Target who needs them for enforcing its access control policy. It is clear that an agreement related to attributes and shared resources must exist between the Origin and the Target in order to establish the access control policies.

2.5.4 WS-Federation

Web-Federation is an important component within the secure framework architecture for Web Services. As we know, Web Services support communication between web applications located in different organizations, and it allows the integration of applications in heterogeneous environment. Web Services base its operation on the Service Oriented Architecture (SOA). Under this context, in 2002, IBM and Microsoft together with other companies defined a reference model to provide security to Web Services from a technological point of view as well as business activity policy (IBM, 2002).

Architecture

Figure 2.18 shows the security architecture model for Web Services.

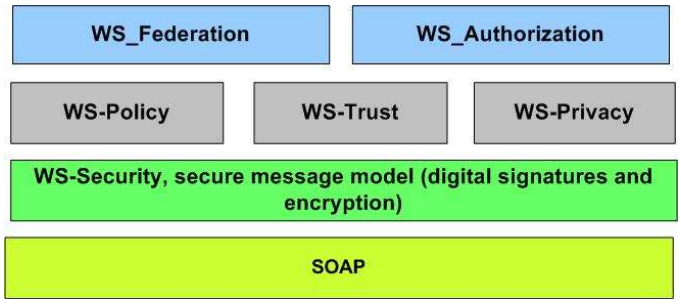


Figure 2.18 Secure architecture for Web Services

The security architecture for Web Services operates with the message transfer protocol of the Simple Object Access Protocol (SOAP), the set of WS-Security definitions extends the functionality of SOAP to include security tokens within a SOAP message. In addition, it guarantees the integrity and confidentiality of the messages by means of the XML encryption and digital signature. The second level of specification (WS-Policy, WS-Trust and WS-Privacy) provides a framework to establish capacities and restriction policies, models of confidence and privacy preferences respectively. Finally, the WS-Authorization and WS-Federation specifications, define the elements necessary to build a Federated Identity Architecture (IBM, 2002).

Elements and operation

The WS-Federation model includes three elements: the Requestor (RQ), that is, an application requiring access to Web Services, the Identity Provider (IdP) or Security Token Server (STS) whose function is to carry out the authentication process and to transmit security tokens with relevant attributes; and the Resource Provider (RP) that provides the resource required by the Requestor (Kaler & Nadalin, 2003). Figure 2.19 shows the interaction between the different components of the architecture based on Web Services.

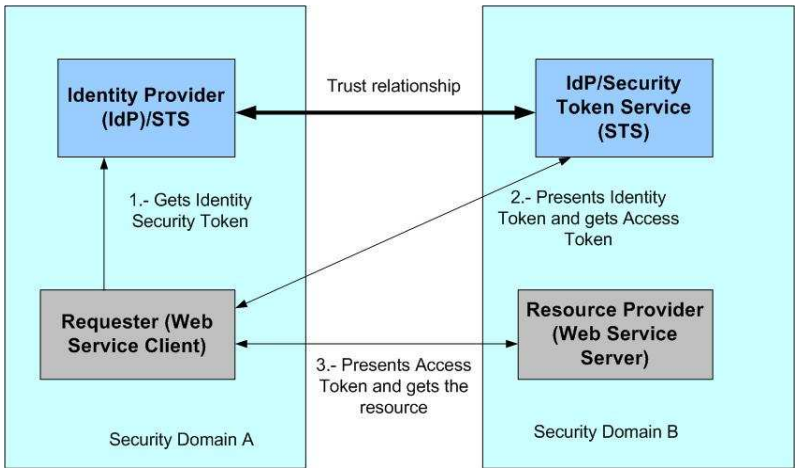


Figure 2.19 Relationship between the components of the Web Services architecture

When RQ in security domain A requests a web service located in another security domain (B in the figure), it is first authenticated by its IdP and obtains a security token with its identity information (1). Depending on the requested web service, an additional access token may be obtained from the other STS in security domain B with the necessary attributes to request the resource (2). Finally, the security token is presented to the Web Service (RP), who evaluates the security token and then applies its access control policy in order to grant access to the protected resource (3).

2.5.5 Initiatives comparison

The two Federated Identity Architectures based on SAML and the WS-Federation initiatives presented in this chapter have similarities and differences, as well as advantages and disadvantages depending on the context and usage cases. In the following paragraphs, a comparison is given in terms of main functionalities:

- **Approach.**- Liberty Alliance and WS-Federated are targeting business interactions whereas Shibboleth focuses on digital academic resources sharing.
- **Identity information storage**.- Shibboleth is based on the centralized model where the identity information is centrally located and only attributes are sent to service providers. Liberty Alliance and WS-Federation, on the other hand, allow that the identity information could be distributed and federated in such a way that the authentication process could be done in any IdP within the Circle of Trust for a particular user.
- **Personal information privacy**.- Shibboleth supports from its origin the management of attributes through its Attribute Release Policies (ARP). In Liberty Alliance and WS-Federated architectures, attributes are divulged under the organization (IdP) control with little or no control from users, some facilities to communicate privacy policies are provided.

- **SSO and web applications**.- All the initiatives support SSO for web applications; however, Shibboleth only supports access to web applications from web browsers, whereas, WS-Federation is only designed for Web Services. Liberty Alliance supports both types of access.
- **Scalability**.- WS-Federation might support a great number of users, IdPs and SPs. This is due to the flexibility of Web Services that may be easily programmed to behave as IdP or SP, and also the their capacity to expand into big and complex structures. With Shibboleth and Liberty Alliance, the roles of the IdPs and SPs are well defined but the need for establishing a secure technological infrastructure and business agreement between the IdP and SP does not offer enough flexibility for building a big CoT.
- **User's security**.- All the architectures are based on standards where the communication channels are encrypted and authenticated, thus guaranteeing a high level of security. However, the main problem is the identity theft which strongly depends on the security controls enforced at the user terminal. Some efforts within the initiatives are currently initiated.

2.5.6 FIA challenges

Despite some important advances carried out in this field, Federated Identity Architectures still face common challenges that represent very important issues for their real implementation. Some of such challenges can be mentioned:

- **Identity theft**.- The theft of an identity represents one of the main issues because generally it remains undetected until the damage is done. In most of the cases, the identity theft does not occur over the communication channels, nor in the IdP repository. It mostly occurs at the user terminal due to the lack of security mechanisms. Therefore research efforts must be allowed to improve robustness and security of terminals.
- **PKI[1] integration.-** PKIs are today largely implemented within companies to support every day enterprise transactions. One important challenge today for the FIAs is to provide integration with PKI so as to extend their functionalities in a transparent way.
- **AAA integration**.- Operators use AAA protocols (e.g. RADIUS, Diameter) to authenticate, to authorize users accessing their networks, and to perform communication accounting. With their ability to authenticate users, and their large geographical coverage, they might serve as IdPs for any applications, and offer this extra identity management service to their subscribers. Moreover, operators are today used to operate inter domain AAA procedures, so that FIA might be naturally deployed over such AAA architecture. Investigations on possible integration of AAA and FIA architectures are clearly needed.
- **P2P[2] application support**.- Use of P2P applications has recently increased very fast. FIA introduction into P2P environment could bring security and a clean identification of P2P entities. However, integration is difficult today as FIA initiatives are based on a client/server model. The exchange of identity information in a P2P federated environment represents an important issue that must be fulfilled.
- **Privacy guaranty and legal compliance**.- In some countries, laws do protect personal information against bad intended use. The current FIA initiatives have some mechanisms that may protect personal information of the user. However, such mechanisms do not fulfill satisfactorily privacy requirements imposed by

[1] Public Key Infrastructure
[2] Peer To Peer

legal frameworks and business agreements. There are some technological initiatives that intend to improve privacy, such is the case of P3P (Privacy Preference Project) which is proposed by W3C[3] to define a standard for web sites to communicate about their practices in terms of personal information collection, use, distribution and laws compliance (Garfinkel & Faith, 2002).

Privacy is really an important aspect in this global technological environment, mainly within the Federated Identity Architecture where exchange of identity information among different organizations is a fundamental service.

2.6 User-Centric Identity Management System

As mentioned in section 2.2.4, user-centric IMS was proposed to provide users more control on their digital identities. This architecture assumes that multiple identities will exist, each of which expressed in a different way and providing different information depending on the context where they are used. Digital identities are represented by elements known as security tokens (ST). A security token is a set of one or more claims expressing information about digital identity. An ST can be a simple text string representing a username, a X.509 digital certificate, a Kerberos ticket or a SAML message. Security tokens are not only useful to convey just authentication information, they can carry any type of identity information. The following sections describe two important user-centric initiatives; the first one is a commercial proposal from Microsoft called CardSpace and the second one is an initiative named Higgins from the open source Eclipse project.

2.6.1 CardSpace architecture

Microsoft has proposed an identity management metasystem called CardSpace, originally Infocard (Chappell, 2006). The objective of CardSpace is to provide a consistent way to work with multiple digital identities, regardless of the kinds of security tokens used. As this technology is centered in the user, it provides an easy to use interface where users can manage intuitively their digital identities; such identities are represented as cards. CardSpace hides the beneath technological aspects and provides users the control of their personal information. Figure 2.20 shows the main components of the CardSpace metasystem; the IdP (Security Token Server in the Microsoft literature) is the entity responsible for storing digital identities and generating the identity cards, the SP (Relying Party in the Microsoft literature) provides services based on the security tokens received, the identity selector is the component at the user side that allows users to manage the identity cards and select the appropriate IdP.

[3] World Wide Web Consortium

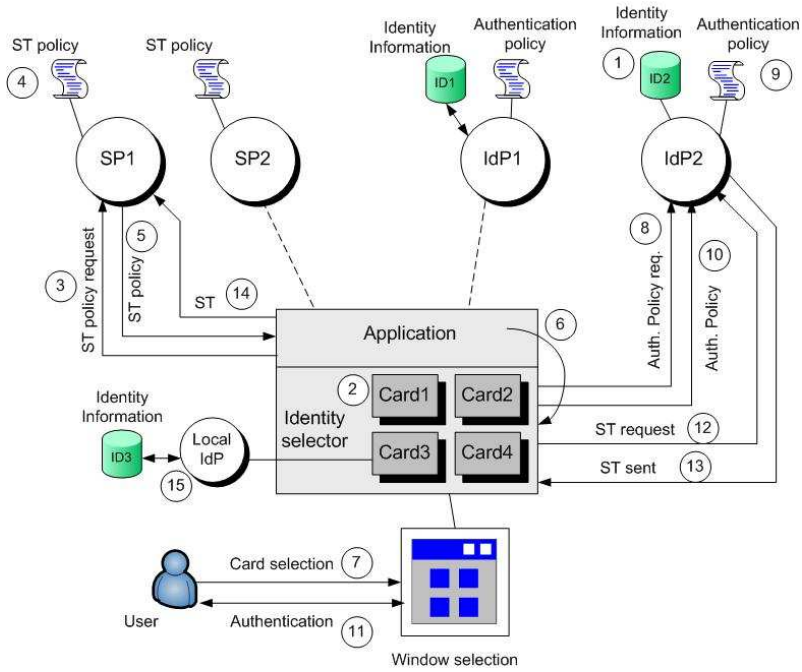


Figure 2.20 CardSpace components and functionality

As we can see in Figure 2.20, the IdP stores the information of the digital identity (1) and creates the cards that are placed in the identity selector (2). When the user accesses a SP, the user application requests the security token policy from the SP (3) which describes the ST format and the required claims (4). The security token policy is sent to the user application (5) and then the identity selector searches for the cards that match the ST policy (6). The cards options are presented to the user, which in turn selects the most convenient card (7). The identity selector requests the authentication policy (8) to the corresponding IdP which describes the authentication mechanism accepted by the IdP in order to create the ST (9). The authentication policy is sent to the user (10) and the user authenticates to use the card (11). Once authenticated, the ST is requested (12) and the IdP creates and sends the encrypted security token to the identity selector (13). At this point, the user can optionally verify the information before sending to the SP (14). CardSpace allows the user to deploy a personal IdP within the local machine in order to provide the digital information directly to the SP without depending on external IdP(15).

CardSpace can be used to carry out the authentication process, but also to send at any time identity information as requested by the SP to provide service personalization. As all the information passes through the user via the identity selector, the user has the total control of which information is released to which SP.

The protocols used to define policies and to communicate the components within the CardSpace metasystem are open and published protocols like HTML, XML, SOAP, HTTP and HTTPS. However Microsoft widely recommends the conformance with the following specifications: WS-SecurityPolicy to define security token and authentication policies, WS-MetadataExchange to exchange the policies, WS-Security to request the ST to IdP and WS-Trust to present the ST to the SP.

Card content

The card in the CardSapce metasystem represents the digital identity and basically is an XML document stored in the local machine. The card does not store any personal information and it is digitally signed by the IdP that issued it (includes the IdP certificate). Figure 2.21shows the structure of the card:

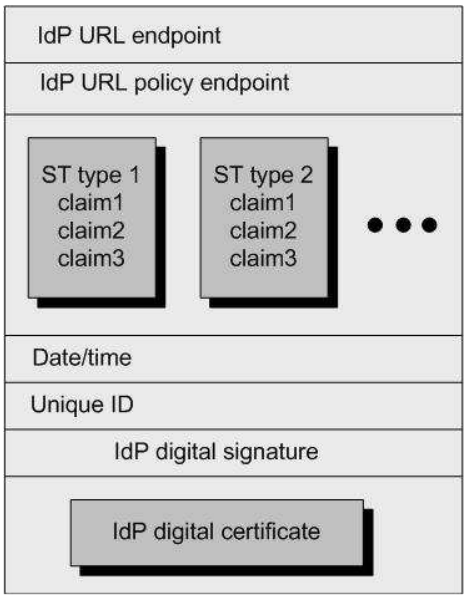


Figure 2.21 Card content

The IdP URL endpoint indicates where to get the security token, meanwhile, the IdP URL policy endpoint provides the address from where to get the authentication policy required to release the security token. The card also contains the different types of security tokens that the IdP is able to provide as well as the claims available for each ST. Additionally, it has the date and time when the card was created and a unique identifier (ID) to distinguish each card. Finally, for security reasons, the card is digitally signed by the IdP and its digital certificate is added.

2.6.2 Higgins

Higgins is an open source identity framework being developed at the Eclipse Foundation project (Eclipse, 2008) that enables users and applications to integrate identity, profile, and social relationship information across multiple data sources and protocols.

Since the release of version 1.0 of Higgins in February 2008, many commercial products based on Higgins 1.0 have been announced by Novell, CA and IBM. The next version (Higgins 1.1) is planned to be released in June of 2009 (Higgins, 2008).

Higgins follows the User-Centric identity management model that has as main original goals:

- To provide a secure and consistent authentication experience to the user based on identity cards (I-cards).
- To build a trusted infrastructure that allows people to selectively share personal information while protecting their privacy.
- To define several provider plug-ins that allows developers to create adapter to legacy systems, protocols and formats types in order to integrate identity systems.
- To provide plug-in adapters to support diverse data sources.
- To organize relationships into a set of different contexts within which a person expresses different roles.

Higgins architecture and operation

As shown Figure 2.22, Higgins is a three layer architecture where the lower component is the Identity Attribute Service (IdAS) layer that provides interoperability and portability across silos of identity data. The intermediate component is the Identity Service (IdS) layer that supports different types of I-cards. The upper component contains the user applications along with the operational elements of the architecture: Identity Selectors, Identity Providers and Relaying Parties (RPs).

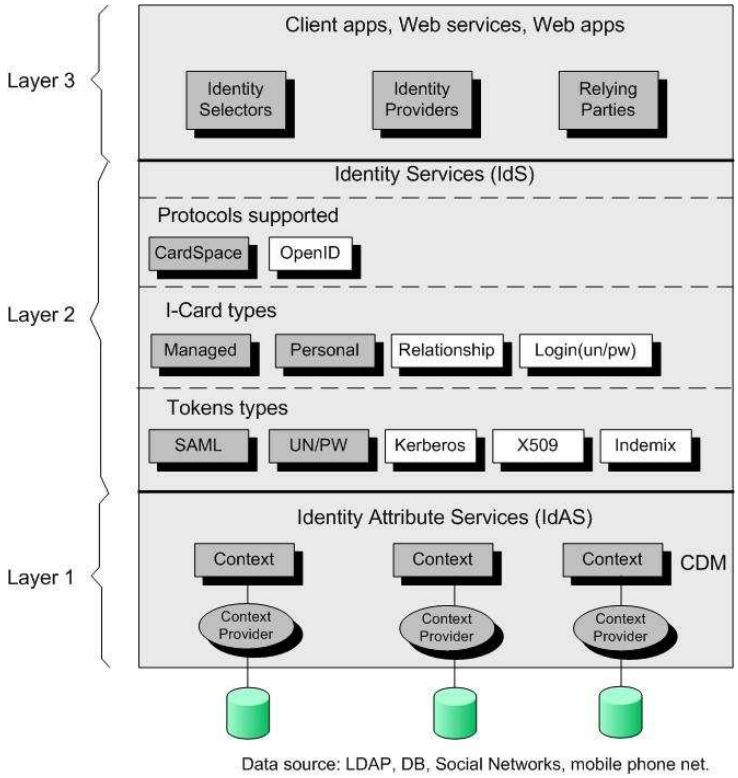


Figure 2.22 Higgins Architecture

IdAS uses Context Data Model (CDM) to provide data abstraction to allow portability across heterogeneous data sources. The data model is expressed by *Contexts* which are a set of entities representing people, groups, organizations, objects, etc. A Context is described by ontologies using standards languages such as Resource Description Framework (RDF) and Ontology Web Language (OWL). Each entity within a Context has attributes with single or structured values. The Context Provider maps the data source to the corresponding Context.

IdS is the layer responsible for handling different security tokens; Higgins 1.0 supports SAML and UserName/Password (UN/PW) security tokens, meanwhile, the next version (Higgins 1.1) will support additional security tokens such as Kerberos, X.509 and Indemix. An I-Card is a piece of identity that provides any kind of personal information. The I-Cards supported initially by Higgins 1.0 are *Managed* (provided by an organization) and *Personal* (generated by the user). The version 1.1 will support *Relationship* and *Login* I-Cards, the first one allows organizing relationships into a set of different contexts, while the second one is to log in with UserName and Password credentials. At protocol level, Higgins 1.0 is compatible with CardSpace of Microsoft and the next version will support OpenID I-Card protocol.

The upper layer has to deal with applications and functionality of Higgins. The applications compatible with Higgins could be client/server, web or web services. Identity selectors are user-centric applications for creating, selecting, sharing and managing I-Cards that represents identity in different contexts and relationships. The Id selectors post a security token that is validated by the Relaying Party. Identity providers generate security tokens which contain

personal information; such information is selectively released by the user to the Relaying Party via the Identity Selector.

2.7 Conclusions

Digital identity management became a relevant security subject of importance due to the great amount and complexity of on line services that the user must interact with. Digital identity information must be exchanged between different organizations in a secure way for preserving personal information integrity and confidentiality. Many IMS models have been proposed, however, the Federated Identity Architecture and the User-Centric model fulfill the requirements of actual online interactions.

FIA involves a set of technological solutions, as well as business agreements between organizations to conform to a trust structure that ensures the exchange of identity information.

The most important initiative of FIA at the moment is Liberty Alliance, since its definitions include not only technological aspects, but also definitions related to business agreements in order to establish a Circle of Trust. This solution is focused on companies to strengthen B2B and B2C relations. Additionally, Liberty Alliance defines a complete framework to incorporate secure identity information exchange based on Web Services.

The Shibboleth proposal is an academic approach where the main objective is to share digital resources between institutions without having to explicitly know the user identity, that is to say, it is an architecture where the privacy of the personal information is widely guaranteed and where most of the accessed resources are under an anonymous basis. Shibboleth is a framework simpler than Liberty Alliance, but it only solves a specific problem of collaboration and resource sharing between academic institutions.

The WS-Federation initiative proposed by Microsoft and IBM focuses basically on the Web Services environment, taking advantage of the impulse made by the Service Oriented Architecture (SOA), which establishes an atmosphere of applications integration between organizations with heterogeneous infrastructures, WS-Federation adds security functionalities and allows the secure exchange of identity information of Web Services.

User-centric is proposed as an architecture that allows users to take control on how the personal information is exposed. This model enhances the privacy aspects of the IMS. However, the user-centric model solves only part of the problem, because, once the identity information is released, the user has no more control of his personal information. Therefore, a detailed analysis of privacy aspects within the IMSs must be done in order to propose a general privacy framework that improves privacy in a complex and online world.

2.8 Conclusions (en français)

La gestion d'identité numérique est devenue un sujet de sécurité d'importance significative du fait de la grande quantité et complexité des services en ligne avec lesquels l'utilisateur doit interagir. Les identités numériques doivent être échangées entre différents organismes d'une manière sûre pour préserver l'intégrité et la confidentialité des informations personnelles. Plusieurs modèles d'IMS ont été proposés. Cependant, l'architecture d'identité fédérée et le modèle d' « utilisateur au centre » (user-centric) remplissent les exigences des services interactifs en ligne actuels.

La FIA implique un ensemble de solutions technologiques, comme des accords commerciaux entre organismes afin de se conformer à un cercle de confiance dans lequel sont assurés des échanges d'informations d'identité.

L'initiative la plus importante de FIA est actuellement Liberty Alliance, puisque ce dernier inclut aussi bien les aspects technologiques, que des définitions liées aux accords commerciaux afin de mettre en place un cercle de confiance. Cette solution intéresse les entreprises car elle permet de renforcer les relations B2B et B2C. De plus, Liberty Alliance définit un cadre complet pour incorporer l'échange sûr d''informations d'identité basé sur des *Web Services*.

La proposition de Shibboleth est une approche académique où l'objectif principal est de partager des ressources numériques entre établissements sans explicitement connaître l'identité d'utilisateur. C'est-à-dire, c'est une architecture où le respect de la vie privée des informations personnelles est largement garanti et où la plupart des ressources accédées se font sur une base anonyme. Shibboleth offre un cadre plus simple que Liberty Alliance, mais il résout seulement un problème spécifique de collaboration et de partage de ressources entre organisations académiques.

L'initiative de WS-Federation proposée par Microsoft et IBM se concentre fondamentalement sur l'environnement de Web Services, il prend avantage du succès de l'architecture Service Oriented Architecture (SOA), qui offre un environnement d'intégration d'applications entre organismes dotés d'infrastructures hétérogènes. WS-Fédération ajoute des fonctionnalités de sécurité et permet l'échange sûr d'informations d'identité de Web Services.

Le modèle d'utilisateur au centre est vu comme une architecture permettant aux utilisateurs d'avoir le contrôle sur la divulgation de leurs informations personnelles. Ce modèle permet de garantir un meilleur respect de la vie privée de l'IMS. Cependant, le modèle de l'utilisateur au centre résout seulement une partie du problème, parce qu'une fois l'information d'identité divulguée, l'utilisateur n'a pas plus aucun contrôle sur ses informations personnelles. Par conséquent, une analyse plus détaillée des aspects liés au respect de la vie privée dans l'IMS doit être faite afin de proposer un cadre général qui apporte une amélioration du respect de la vie privée dans un monde complexe et en ligne.

Chapter 3

Privacy overview

Privacy is a term difficult to define because its meaning depends on the context and the geographical location. However, in most of the countries, privacy is recognized as a fundamental right in all major international treaties and agreements on human rights and in the constitutions of most countries in the world. The recognition of privacy dated from many centuries, but the most important at an international level can be found in the 1948 *Universal Declaration of Human Rights*, specifically in the article 12 that states: "No one should be subjected to arbitrary interference with his privacy, family, home or correspondence, nor attacks on his honor or reputation. Everyone has the right to the protection of the law against such interferences or attacks" (Privacy, 2006). At regional level, some treaties regarding privacy followed, like the article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms and the article 11 of the American Convention on Human Rights (Dumortier & Goemans, 2004). Basically, most countries have developed any kind of legal framework regarding privacy.

The fast development of the information technology mainly from the beginning of the seventies, made insufficient the basic recognition of privacy in the initial treaties. That demands new legislation at international, regional as well as local level to recognize privacy as the protection of personal digital information from collection to storage and dissemination. The two most important international initiatives in 1980 and 1981 respectively were the OECD "Guidelines on the Protection of Privacy and Transborder Flows of Personal Data" (OECD, 1980) and the Council of Europe "Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data" (Europe, 1981). Many other laws and recommendations emerged in different geographical regions and countries, but most of them are based on the principles stated by the OECD and Council of Europe initiatives.

We have seen in recent years that many digital services require the collection of personal information in order to give the subject a better and personalized service. (Landesberg & al, 1998) showed at the end of 90´s that more than 90% of the web sites collect personal information with few or non awareness regarding privacy, however, nowadays enterprises and users are considering privacy issue an important element for the increasing and consolidation of online services.

This chapter describes the main legal frameworks regarding privacy of personal data developed by the OECD, European Union, United States and in particular by the Mexican government. Besides privacy legal frameworks, technology plays an important role regarding privacy. There are technologies that could compromise privacy if they are not managed properly, those technologies are known as Privacy Intrusion Technologies or PITs. On the other hand, there are technologies which improve or enhance privacy, those technologies are known

as Privacy Enhancing Technologies or PETs. In this chapter, some examples of both technologies will be presented, emphasizing how PETs could be related with the privacy principles of the OECD guideline. Finally, privacy policy languages are presented as an important PET technology for the privacy model developed in the following sections.

3.1 Privacy Legal Frameworks

The following sections will explain the main privacy principles of the OECD, as well as the main legal framework regarding privacy in Europe, USA and Mexico.

3.1.1 OECD Guidelines

The Organization for Economics Co-operation and Development (OECD) is an organization formed by 30 members that share a commitment to democratic government and market economy. The OECD was the first international organization to make an attempt to unify the initiatives relative to the protection of personal information. In 1980, the OECD issues the *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (OECD, 1980).

We can summarize the following eight principles taken from the OECD Guidelines that are the basis for the development of the majority of the initiatives mentioned:

1. **Collection Limitation**.- The amount of personal data should be limited and obtained by lawful and fair means.
2. **Data Quality**.- The personal data should be relevant to the purpose they are to be used and such data should be accurate, complete and up to date.
3. **Purpose Specification**.- The purpose of the data collection should be specified on time, as well as any change of purpose.
4. **Use Limitation**.- Personal data should not be disclosed or used for other purpose than stated by the Purpose Specification.
5. **Security Safeguards**.- Personal data should be protected by security safeguards against risks that its confidentiality and integrity are compromised.
6. **Openness.**- Personal information policies and practices should be notified to the subject from whom information is collected.
7. **Individual Participation**.- Subjects must be able to access its own personal information, and be able to challenge the accuracy and completeness of it.
8. **Accountability**.- The data controller (who is competent to decide about the content and use of personal data) should be accountable for complying with the previous principles.

The Privacy Guidelines represent international consensus on general guidance concerning the collection and management of personal information. The privacy principles defined are characterized by their clarity and flexibility to adapt to technological changes. The principles are applicable at both national and international levels.

3.1.2 European Union Directives

In 1981, the Council of Europe adopted the *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data* (Europe, 1981) known as the "Council of Europe Convention 108". This convention is the first binding international instrument which protects the individual against abuses. It may accompany the collection and processing of personal data and seeks to regulate at the same time the transborder flow of personal data.

This convention is legally binding on member and requires them to enact personal information protection legislation that will cover the use of such information in the public and private sector.

From 1981 until 1990, the EU released some recommendations regarding personal information protection focused on a specific application such as: recommendation 81 regarding automatization of medical database, recommendation 83 related to the protection of data used to support scientific research and statistics, recommendation 85 to protect personal information used in direct marketing and the recommendation 90 regarding the protection of payment data (Guerrier, 2008).

In 1995, the European Union adopted *The Directive (95/46EC) on the Protection of Individuals with regard to the Processing of Personal Data and the Free Movement of Such Data* (Council, 1995), known as "The EU Data Protection Directive", this proposal was the response to the perception that were in inadequacies with the OECD Privacy Guidelines and the Council of Europe Convention 108. The two overall objects of the directive are the protection of information privacy by member states of the EU, and the prevention of restrictions on free flow of personal information between EU members states for reasons of privacy protection. The EU Data Protection Directive sets out nine principles mainly based on the OECD Privacy Guidelines, but affording a high level of protection. Under the EU directives, every EU country must have a data protection commissioner or agency that enforces the rules. It is expected that the countries with which Europe does business, will need to provide a similar level of regulation.

3.1.3 USA Initiatives

The implementation of the EU Privacy Directive would prohibit the transfer of personal data to non-European Union nations that do not meet the adequate standard for privacy protection. This has brought serious challenges for the United States since from the EU opinion, the United States do not meet the directive´s standards for the protection of privacy and that limited the opportunity to U.S. companies of doing business with EU members (Sun, 2003). The American approach to privacy protection is driven by business interest and uses a sectoral approach that relies on a mix of legislation, regulation, and self regulation, as compared to the EU´s rights based approach. These different privacy approaches, may significantly disrupt trans-Atlantic trade, as well as impede the development of e-commerce. Since the adoption of the Directive in 1998 by the EU, the U.S. Department of Commerce started an intense negotiation with the European Commission in order to resolve their privacy policy discrepancies. The result of such negotiations was the development of the *Safe Harbor framework* (Commerce, 1998) that enables the U.S. organizations to provide equivalent privacy protection, as defined by the EU Directive. The U.S. companies could optionally self-certify annually to the Department of Commerce by writing that it agrees to adhere to the Safe Harbor requirements, which include seven privacy principles. Safe Harbor implementation is at least not as effective as it was expected in 2000 during its implementation. While the EU continues to enhance privacy protection, in the United States, the Congress adopted anti-terrorism measures after September 11, 2001 that created holes in the privacy of personal data.

An example of privacy initiative focused by sector is the *Health Insurance Portability and Accountability Act of 1996* - HIPAA (Health, 1996) established by the U.S. Department of Health and Human Services, such Act addresses the use and disclosure of individuals´ health information. The major goal of the privacy rule is to assure that individuals´ health information is properly protected while high quality health care is provided. Another specific legal framework is issued by the U.S. Federal Trade Commission (Federal, 1974) regarding the protection of personal data of consumers that deal with businesses, as well as financial companies.

3.1.4 Mexican Directives

The Mexican Constitution, in accordance with the Universal Declaration on Human Rights, the American Convention on Human Rights, and the Declaration of Principals of the World Summit of Information Society acknowledges the protection for the individual's right to privacy

and data protection as one of the most important fundamental rights (Andrews, 2003). In 2005, significant regulatory advances occurred in Mexico with the enactment of the *Directives for Personal Data Protection* (IFAI, 2005) that protects personal data processed, stored and shared by Federal Government agencies. Besides, there are diverse sector laws that protect privacy and personal data such as; Federal Consumer Protection Act, Geographic and Statistics Information Law, Law for Regulating Credit Information and Commercial and Civil Code. The directives establish general policies and procedures that Federal Government agencies must comply to protect privacy of personal information. The directives are influenced by the OECD privacy framework and the European data protection directives, and contain the eight privacy principles. Regarding personal data protection for the private sector, the Mexican congress is discussing a constitutional reform and a widely national act to protect personal data in the private sector as well.

3.1.5 Summarize of the Legal privacy initiatives

There is a general consensus regarding the importance of the personal information privacy. The initial efforts started at beginning of the seventies with the accelerated development of information technology. The OECD was the first international organization that issued general guidelines containing basic privacy principles that were taken as basis for the development of other legal frameworks. The European Union improved the OECD principles and issued their Directives that protected the personal information between the EU members and at the same time facilitated the free flow of personal information among the EU member´s countries.

Despite their excellent structure, implementation and enforcement, the EU Directives imposed trade restrictions with countries outside the EU that has weak or null legal legislation regarding privacy protection. That was the case of United Stated who had to develop in coordination with European Commission a legal framework called Safe Harbor in order to be compatible with the EU privacy Directives, so that, U.S. organizations could exchange personal information with member countries of EU. The United State approach to privacy has been by sectors; as an example, we can mention the Federal Trade Commission initiatives for commercial issues or the HIPPA for health information.

In the other hand, the terrorism events of September 11 2001, has promoted the development of anti-terrorism initiatives, as the USPatriot Act, that goes against the privacy of personal information, since it allows the U.S. government to access any personal information that could compromise national security. There are additional efforts at international, regional or national levels regarding the development of privacy legal frameworks as the case of Mexico and other countries around the world. Table 3.1 shows the initiatives previously described, from the point of view of origin, focus and scope.

Legal Privacy Framework	Origin	Focus	Scope
OECD Guidelines	OECD	General	International
EU Directives	European Union	General	EU Members
Safe Harbor	U.S. and EU	Trade	U.S. organizations and EU countries
Federal Trade Commission Act	U.S.	Commercial	U.S. enterprises and consumers
HIPPA	U.S.	Health	U.S. citizens
Personal Data Protection Directives	Mexico	Government	Government Secretaries and citizens

Table 3.1 Privacy legal frameworks comparison

Privacy of personal information can be protected by a legal framework as mentioned in previous sections. The regulation can be deployed and enforced by government or internally by local authorities, however, the rapid rise of interconnected global network and the increasing flow of personal data across national borders have raised awareness about privacy concerns. In that context, the OECD has recognized the role that technology can play in enhancing privacy in the online environment. The OECD has issued additional privacy guidelines (the 1998 Ministerial Declaration) focusing on the Internet and global electronic commerce.

3.2 Privacy Intrusion Technologies and Privacy Enhancing Technologies

From the technological point of view, there are some current technologies that represent a serious risk to privacy; they are named Privacy Intrusion Technologies (PITs). However, there are technological countermeasures known as Privacy Enhanced Technologies (PETs). The main PITs and PETs technologies are briefly explained in the following sections, along with how they relate to the privacy principles previously mentioned.

3.2.1 Privacy Intrusion Technologies

Digital identities as part of personal information are sensitive to threats that compromise their privacy. Those threats can be classified as technological and non-technological, the latter ones are related to physical events such as physical information theft, natural disaster or social engineering. These threats are out of scope of the present work. With respect to technological threats, we will focus on current technologies that could represent a serious risk to personal information privacy. Such technologies are classified into the application areas (Privacy, 2006) described in the following sections.

Identification Systems

Identification systems are those systems and technologies used to identify a person within an application scope. One of the most frequently elements used in such systems is the ID card which can be used for specific applications such as credit card, healthcare card, driver license card, passport, among others, or it can be a multi-purpose card. ID cards can store much personal information that can be exposed or misused during the authentication process.

Additionally, two emerging technologies have been incorporated into ID cards, biometrics and Radio Frequency ID (RFID); they are used to strength and facilitate the automatization of authentication process respectively. However, biometric information is very sensitive because it is tightly bound to the person and RFID is a technology that facilitates the transmission of information in a contactless way. Those technologies could represent a serious risk to personal data privacy if they are not protected properly.

One example of the controversy of the use of these technologies is the U.S government initiative for incorporating without protection such technologies to the electronic passports. Many organizations and countries are against the U.S. initiative, but the State Department argues that the U.S. national security is more important than the personal privacy, mainly after the terrorism events of September 11th.

Surveillance of communications

Electronic surveillance is the technique of intercepting electronic communications that is used in special cases such as crime investigation or technical support. However, it represents a privacy issue due to the facility of gathering personal information without the knowledge and consent of the user. There are many technologies to carry out the surveillance of

communications. One of the most frequently used is the packet *sniffer*, which is an equipment connected directly to the communication network, so it can "listen" all the traffic flowing through it. The *sniffer* is used mainly to debug network problems, but it could intercept personal information mainly when the information is not encrypted.

Proxy servers also represent privacy risks. A proxy server is a device that interconnects two or more communication networks at application level mainly used to improve functionalities; therefore, all information exchanged between applications passes through the proxy server. In that sense, personal information could be easily collected.

When the technologies above mentioned are combined with mobile technologies such as cellular, Wi-Fi or GPS (Global Positioning Satellite) systems, not only the personal information could be intercepted, but also the physical location of users can be determined precisely.

Web browsing

One of the most used applications in on-line transactions is the Web Service, supported by the HTTP protocol. That protocol along with the identifiable IP address, enable to trace user information related to his geographical location as well as the characteristics of his web browser. Although the facility of tracing the user represents a privacy risk, the real threat to privacy with Web Service is the use of cookies.

Cookies are defined as a technological means for gathering information in order to facilitate the Web browsing. The information collected by cookies generally is related with personal information such as name, e-mail, address, among others, as well as personal preferences. The problem with cookies is that the user has little or no control on the information collected by cookies and the way such information is going to be used. The last point goes against one of the fundamental privacy principles previously mentioned.

An additional problem with the Web is the facility to integrate embedded software within the web-based applications such as JavaScript or Active X; they permit to run applications on a client's PC, and therefore, to gain access to the user personal computing environment and the data held within it. Cookies and embedded software are techniques used to build profile databases that contain the preferences, activities and characteristics of users; this practice is called *E-profiling* and serves to send users commercial advertisements without their consent nor their knowledge.

3.2.2 Privacy Enhancing Technologies

PETs refer to a wide range of technologies that help protect personal privacy. They can empower users to control the disclosure, use and distribution of online personal information. They can also aid organizations in enforcing their own privacy policies and practices (OECD2, 2003).

User side technology

These technologies are deployed directly at the user side. One of the most used technologies is the data encryption for storing and transmitting personal information. The encryption mechanism guarantees to some extent the confidentiality of personal information when it is transmitted or stored. In some countries where encryption algorithms are controlled and all information must be sent without encryption, users may use an alternate technology such as steganography in order to keep privacy, this technology consists in hiding information within a normal file, normally a graphic file. The idea is to add information in unused bits of the file, so that injected information remains undetectable. There are additional technologies that can be used to enhance privacy, for example, tools that help managing cookies by blocking, selecting, deleting or viewing them. Such functionalities are being incorporated in recent versions of web browsers. Privacy languages may allow at the user side to define his privacy preferences in

order to build the privacy policies that will protect his personal information (Senicar & al, 2003).

Third party mechanisms

These types of privacy mechanisms are deployed between any two entities exchanging personal information. Examples of such technologies are among others: pseudo-identities and anonymizers, access control and privacy languages. The mechanisms to build pseudonymity and anonymity are also known as identity protectors. An identity protector can be seen as an intermediary between the user and the Service Provider (SP); it provides the functionality of hiding the real identity for those services where the knowledge of the real identity is not necessary. Access control mechanisms grant or deny access to personal information based on the request, the environmental context and the privacy policies associated to such personal information (Senicar & al, 2003).

3.2.3 Synthesis of PETs and their relationships with privacy principles

Table 3.2 explains with more detail some of the PETs above introduced and shows how they could support the privacy principles given by OECD.

Technology	Description	Supported OECD privacy principles
Encryption	One of the oldest technologies to preserve the confidentiality of data to be transmitted or stored has been encryption. The encryption mechanism consists of data transformation so any unauthorized person can not have access to it. Nowadays, there are many algorithms to carry out data encryption. Some mechanisms are based on symmetric key systems where the same key is used to encrypt and decrypt. Asymmetric key systems might also be used. They refer to a pair of keys that are different but correlated, and that can be managed thanks to some PKI (Public Key Infrastructure) architecture.	Security Safeguards
Cookies management	Applications that allow the user to know when cookies (cf. section 2.1) are being written to the hard drive, to manage the acceptance, and to view what information is stored in an individual cookie.	Collection limitation, Security Safeguards
Pseudonymity and Anonymity services	Pseudonymity and Anonymity service is provided by a trusted third party, which allows users to interact with different Service Providers without exposing their real identities. This type of service can be used for several applications such as payment (e-cash), web browsing or email sending.	Collection limitation, Use Limitation
Access control mechanism	This mechanism allows organizations to enforce privacy policies when personal information tries to be accessed.	Use Limitation, Security Safeguards
Privacy preference languages	Languages mainly based on the XML standard and used to allow users to express their privacy preferences while browsing web servers. Similarly, they facilitate organizations to express	Collection Limitation, Purpose Specification, Openness

	the privacy practices within the Web servers.	
Policy enforcing languages	These languages allow organizations to implement access control and policies enforcing mechanisms at the points where personal information is exposed.	Use Limitation, Security Safeguards
Policy negotiation languages	They are used to negotiate a common privacy policy when both, the requester and provider of personal information defined privacy policies.	Purpose Specification, Use Limitation, Openness

Table 3.2 Main Privacy Enhancing Technologies (PETs)

From a technical standpoint, none of the identified PETs uses a full range of functionalities that would make it possible to provide total privacy protection. Users and organizations must therefore combine several tools to ensure a certain level of privacy protection. In general, technologies can be employed to help achieve some of the internationally recognized privacy principles. They can be used in organisations that have chosen either a self-regulatory or legal approach to privacy. Although PETs help protect individuals' privacy, they cannot guarantee the privacy of information once it is given to an organization or business. Another restriction of PETs is that some are not easy to use, so users are not sure to purchase, install and operate them as client-side tools on their computers. It needs to be ensured that there is an understanding of what privacy solutions PETs can provide, as well as an understanding of their limitations in fully addressing all privacy needs.

The following sections will describe with more details the privacy policy languages. They are one of the most important PETs used to guarantee privacy preference and privacy enforcement within an information system where personal information is widely exposed.

3.3 Privacy Policy Languages

Privacy policy languages can help with several of the stages involved in managing privacy policies like writing, combining, enforcing, among others. Some policy languages are designed to help organizations express their privacy policies and to facilitate their enforcement, while other languages are designed to help users define their privacy preferences (Kumararugu & al, 2007).

In this context, we can classify policy languages in three different groups based on their functionality within a privacy management framework:

- Policy languages to express privacy preferences.- This first group of languages allows the users and organizations to express their privacy preferences in a human-readable manner and then to be mapped into a set of rules in machine-readable format. This type of languages is used at the stage when privacy policies are created or modified.
- Policy languages to negotiate privacy policies.- The second group of policy languages is designed to negotiate a common privacy policy when both; the requester and provider of personal information have privacy policies, one expressing the intended use and the other one the desirable use. The applicability of this group of languages is at the transactional stage, before any personal information is exchanged or in each transaction.
- Policy languages to enforce privacy access control. The third group is used to implement the rules which will apply at the access control point in order to enforce the privacy policies. This group of languages is also deployed at the transactional stage.

The following sections will explain with more details each of the privacy group language, giving some description of the most important implementations for each group.

3.3.1 Privacy Preference Languages

As mentioned before, Privacy Preference Languages were designed to allow principals and Personal Information (PI) requesters to express privacy preferences and privacy practices respectively. In this first group we could mention P3P (Platform for Privacy Preferences) (Cranor, 2002), APPEL (A P3P Preference Exchange Language) (Cranor2, 2002) and XPref (X-Path Based Preference Language) (Agrawal & al, 2003).

Platform for Privacy Preference (P3P)

One of the first preference language specification developed at the end of 1990s was the Platform for Privacy Preferences (P3P). It was standardized in 2002 by the World Wide Web Consortium (W3C). P3P provides a standard way for Web sites to communicate about their practices regarding the collection, use, and distribution of personal information. P3P includes a machine-readable privacy policy syntax expressed in XML and a simple protocol that web browsers and other user agents can use to fetch the policy, compare it with local preferences and act accordingly. Although P3P provides a technical mechanism for ensuring that users can be informed about privacy policies before they release personal information, it does not provide a technical mechanism for making sites act according to their policies (Cranor, 2002).

A P3P policy is structured as a sequence of Statements containing the following elements:

- **Purpose**.- Describes purposes for which information is collected. There are 12 purposes already defined in the specification, but new purpose elements can be added.
- **Recipient**.- Describes the intended users of the collected information. Multiple recipients can be specified in one statement. P3P has 6 predefined types of recipients.
- **Retention**.- Defines the duration for which the collected data will be kept. The specification has already defined 5 types of retention.
- **Data-Group**.- Provides the type of data that are collected for stated purposes.
  A P3P policy can specify if the data are optional or not for each attribute to be released.

Figure 3.1 shows how P3P works and the interactions between the user agent and the Web server. Basically, P3P protocol is an extension of HTTP. The user agent fetches the policy reference file which contains the locations of the privacy policies associated to each part of the Web site. The user agent requests the policy, then it is compared with its local privacy preferences (written with any preference language) and the corresponding action is taken. Generally, the action is taken manually by the user. Finally the web page is accessed.

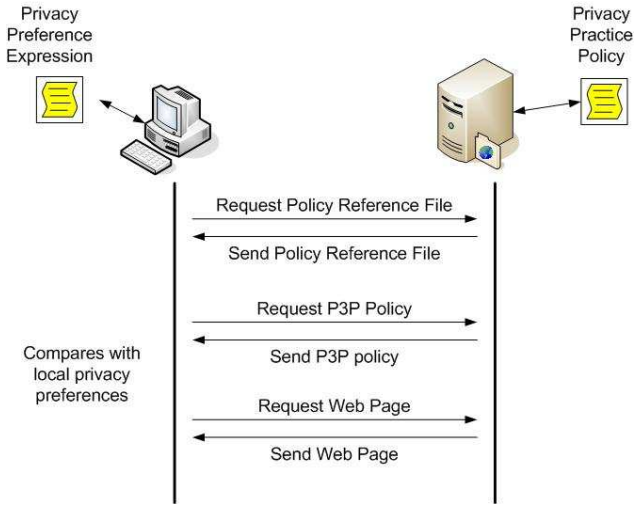


Figure 3.1 P3P operation

There are user agents already built-in the Web browsers or integrated as a plug-in. However, P3P agents can also be built into electronic wallets (PDAs or cellular phones), standalone applications, or any other tool.

The user can use any preference languages to specify his privacy preferences within the agent; the following section will describe a couple of such languages.

Privacy Preference Languages for Users

At the same time that P3P was released to allow organizations to express their privacy practices in the Web site, the W3C also designed APPEL (A P3P Preference Exchange Language) which allows the user to express his privacy preferences, to query the P3P policy, and to make decisions accordingly (Cranor2, 2002).

The privacy preferences expressed in APPEL, are basically a list of rules written in XML format that must match against the P3P policy. These rules consist of two parts:

- **Rule behavior**: specifies the action to be taken if the rule matches. The behavior can be *request* or *block* if the policy conforms or does not conform the preferences.
- **Rule body**: Provides the pattern that is matched against a P3P policy.

As an example, let's assume that the privacy preference specification of a user establishes:

"Block sites if my personal information is used for contact or telemarketing purposes"

The privacy preference expression written in APPEL would look like:

```
<appel: RULESET>
<appel: RULE behavior = "block">
        <PURPOSE connective = "or"
                <contact\>
                <telemarketing\>
        </PURPOSE>
</appel:RULE>
<appel:RULE behavior = "request">
        <PURPOSE
                <otherwise\>
        </PURPOSE>
</appel:RULE>
</appel:RULESET>
```

The first rule blocks explicitly the purposes contact and telemarketing, and the second rule applies for any other purpose.

At first glance, it looks that APPEL is an appropriate language for expressing privacy preferences at user side. However, (Agrawal, 2003) analyzed and concluded that APPEL has serious design problems:

- With APPEL, users can only directly express what is unacceptable. That is, in APPEL expressions, what is not explicitly forbidden is implicitly allowed. To construct expressions with the opposite logic is not possible.
- APPEL does not include logical operator for combining multiple rules in a ruleset; they are evaluated strictly in order.

Those limitations make APPEL a preference language difficult to use even for simple privacy preferences. (Agrawal & al, 2003) proposes a new preferences language called XPref which is an extension of XPATH, and it has the objective to remove the APPEL deficiencies. XPATH is an expression language used to match the structure of an XML document against a

path notation used for navigating through the hierarchical structure of an XML document. Therefore, XPref uses a subset of XPATH elements to express rule conditions that can be matched against P3P policies.

For example, if we want to express an explicit acceptable preference in XPref specifying that only *local-analysis* and *statistics* purposes are acceptable, the privacy preference looks like:

```
<XPref: RULESET>
<XPref: RULE behavior = "request">
        every $name in
        STATEMENT/PURPOSE/*  satisfies
                (name ($name) = "local-analysis" or
                 name ($name) = "statistics")
</XPref:RULE>
<XPref:RULE behavior = "block"  condition="true"/>
</XPref:RULE>
</XPref:RULESET>
```

The expression containing the element "every", is an XPath expression that searches within all the P3P policy structure (even though the P3P could contain multiple statements) for the purposes specified, contrary to APPEL that would match with the first statement without searching in all the entire policy structure.

3.3.2 Privacy Policy Negotiation Languages

In this document we present two approaches of policy negotiation: WSPL (Web Service Policy Language) and Liberty Multileveled Policy. In the first one, the two policies are merged and a new privacy policy is resulted with rules satisfying both requirements. The second approach is proposed by Liberty Alliance, and basically it defines a set of pre-negotiated privacy policies. The requester of personal information proposes one policy, the provider compares it with its own proposal and then the most restricted policy is selected.

WSPL

WSPL is a policy language that allows Web Services negotiate policies by supporting the merging of policies from the Web Service Consumer (WSC) and Web Service Provider (WSP), its syntax is a subset of XACML (Anderson, 2004).

The policy negotiation can be used to negotiate different aspects such as: quality of service, authentication, authorization, service options, among others. However, in our context, the relevant aspect is the negotiation of privacy policies.

For WSPL, a policy is a sequence of one or more rules, where each rule represents an acceptable choice. The rules are listed in order of preferences. A rule is a sequence of predicates, where a constraint on the value of an attribute is specified.

Policy negotiation may be either static or dynamic. It may be done once for two parties that have static policies, or it may be done at runtime based on policies that represent dynamic constraints.

The policy merging algorithm establishes that the rules are paired in all possible combinations in order to produce a single new rule. If the rules can not be combined, the pairing is eliminated. Predicates that constrain the same attribute must be combined so that the resulting predicate represents their intersection.

As an example, the following sentences describe the policy preferences and the policy intended usage from the attribute provider and requester respectively:

Attribute provider

- My personal address may be released to third party if it is kept for no more than 30 days by the third party.
- My personal address may be kept for one year but not released to third party.

Attribute requester

- I will not release your personal address to any third party.
- I will release your personal address only if it is kept by the third party for no longer than 20 days.

The privacy policies from the requester and provider, as well as the policy resulting from the negotiation are presented in Figure 3.2.

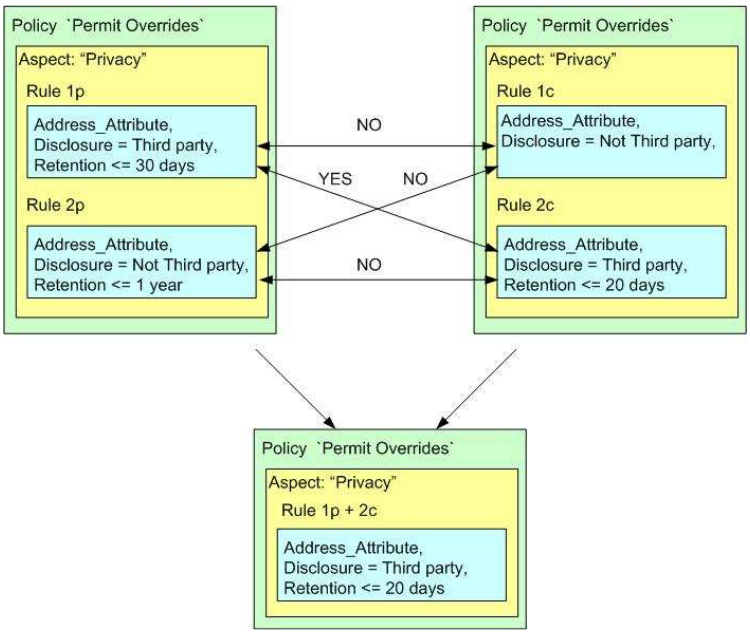


Figure 3.2 Policy negotiation and merging

All the rules from one policy are paired with the rules from the other policy, in this case only the rule 1 of the policy provider can be combined with the rule 2 of the policy requester because the address information can be disclosed to third party, but the time of retention is less or equal to 20 days, resulting in the merging policy that satisfies both policies. Both policies must have the same *Combining Algorithm* for the resulting policy, in this case, it is the *Permit Overrides*. Combining algorithms are explained in section 3.3.7.

Liberty Multi-leveled Policy Approach

The Liberty ID-WSF does not specify a specific negotiation policy language, but it proposes an architecture that incorporates usage directives facility that allows requesters to designate their intended use for requested data, and allows providers to designate the permitted use of released data. The Liberty components must agree in advance on a common set of supported policies and the expression language to use to represent those policies (Landau, 2003).

The WSP (Web Service Provider) acts on behalf of the Principal, meanwhile, the WSC (Web Service Consumer) represents the SP who is demanding an attribute. When the request for personal data from the WSC is made, the request is accompanied by the reference to the

intended privacy policy. If the privacy level of the WSC matches or exceeds the WSP´s, the requested data are disclosed.

The Liberty privacy policy framework proposes a multi-leveled policy approach where the SP/WSC and the Principal/WSP communicates their policies to each other, therefore, an "intersection" must be found between these two policies. To facilitate such intersection, a small number of "standardized" privacy policies can be referred to, usually, the number of privacy policies are related to the number of privacy levels defined for personal information. When the request for personal data from the WSC is made, the request is accompanied by the reference to the privacy policy. If the privacy level of the WSC matches or exceeds the WSP´s policy, the requested data are disclosed.

The WSP must check the following rules when deciding whether to release the requested attributes:

UsagePolicy = The policy chosen by the principal
PrivacyPolicy = The policy used in the request by the SP

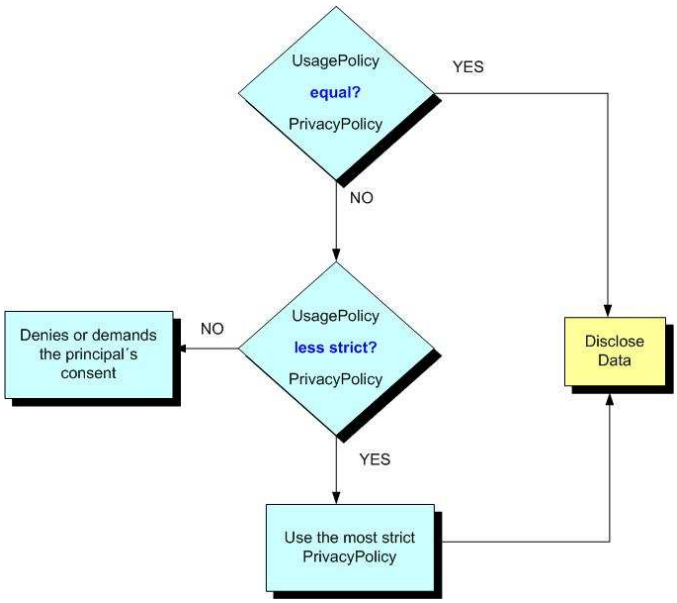


Figure 3.3 Multi-leveled policy matching

Figure 3.3 shows that if both policies are equal, then the personal information is disclosed. If they are different, an additional verification is done to guarantee that the PrivacyPolicy is the strictest. If none of the validations are true, then the access is denied or the consent of the principal is demanded.

Liberty ID-WSF framework enables participants to associate a privacy policy with a message by using SOAP headers. The framework supports policies encoded in any privacy preference language. The choice of privacy preference language and the actual privacy policies for use within a "circle of trust" should be designed with participating service providers, industry norms, and regulatory requirements.

3.3.3 Privacy access control languages

The privacy access control languages allow us to model the authorization policies or privacy policies to an access control model so that such policies could be enforced. First, an overview of the basic access control models is exposed, then the main components and functionally of the access control model is explained.

One of the most complete languages is XACML (eXtensible Access Control Markup Language) (Moses, 2005), which is an extension of XML and was designed mainly to allow access control, but it has been used successfully for privacy purpose.

Access Control Models

An access control model (ACM) allows us to specify the authorization policy associated to any object or information resource, in the case of privacy, such resources concern personal information.

Basically, an authorization policy establishes "which *subject* has the authorization of doing certain *action* over an *object*". For privacy, it is necessary to add some terms such as: the *purpose* of the action, some *obligations* and contextual *conditions* under which the action is taken. Therefore, we can establish an authorization privacy policy as: "which *subject* has the authorization of doing certain *action* for an established *purpose* over an *object* under specific *conditions*, and the *obligations* of the subject once the permission is granted".

From the beginning of the 70`s, many access control models have been proposed (Cuppens, 2006). The most basic ACM is the *Identity Based Access Control* (I-BAC), where the access control is based on the identity of the subject and the identifier of the object. In order to implement this model, it is necessary to make an access control matrix that specifies for each object, who has the authorization of doing certain action. The disadvantage of this model is the complexity for expressing the authorization policies, as well as the difficulty to manage the access control matrix each time a new object or subject is created.

To overcome that problem, new ACMs have been derived from the I-BAC model, basically, structuring the authorization policy around the subject, object or action. Examples of new ACM are: *Role Based Access Control* (R-BAC) which defines the authorization based on the role played by the subject within the organization, *View Based Access Control* (V-BAC) which structures the authorization around views of the objects, this model is widely used in Data Base applications, or *Organizational Based Access Control* (Or-BAC) which centers the authorization on the concepts of organization, within this ACM is taken into account; the role of the subject (usually a group or team), the activities to develop, and the view of the objects.

For simplicity, next sections explain how to use an access control language to implement the I-BAC model with the corresponding extensions of purpose, obligations and conditions in order to enforce privacy policies.

Privacy access control model

The model to implement access control for attributes protected by privacy policies is based on the I-BAC model above described. The functional modules and interactions between them are shown in the following diagram:

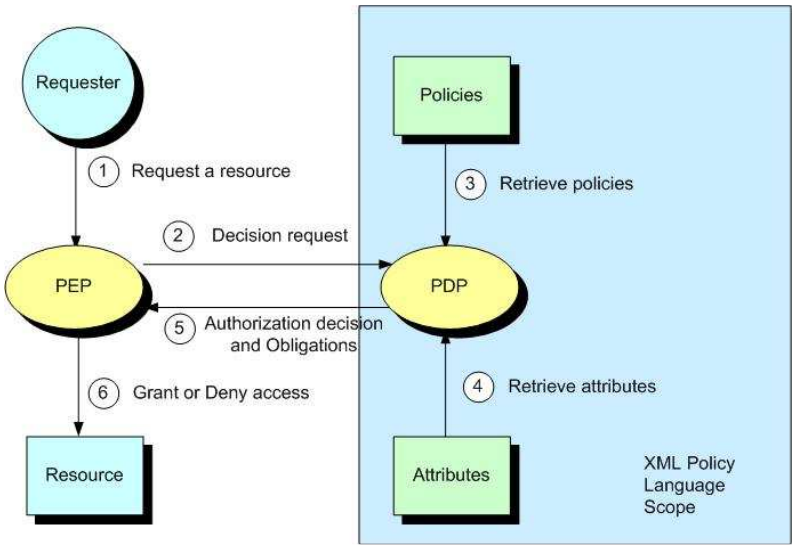


Figure 3.4 Policy Enforcement Model

The Requester represents any entity (user, process, device, etc.) that is able to demand access to a resource. The *Policy Enforcement Point (PEP)* is the functionality that intercepts the request from the *Requester* and allows or denies the access based on policy evaluation. The **PEP** functionality can be located at the *Requester* side, *Resource* side or to be integrated in a third part component, its implementation is local dependent and it is out of the scope of the policy languages specifications. *Policy Decision Point (PDP)* is the component responsible for evaluating the policy, based on the access request parameters and additional attributes. As a result of the evaluation, the **PDP** returns an *Authorization Decision* to the **PEP**, so that it can enforce the policy (Moses, 2005).

As shown in Figure 3.4, the *Requester* requests a service (1), such request is intercepted by the **PEP** and generates a new request for authorization with the parameters sent by the *Requester* and it sends it to the **PDP** (2), the **PDP** receives the request for authorization and it evaluates the corresponding policy (3). The policy to be evaluated could be a single policy or could be the result of combining many distributed or centralized policies. During the policy evaluation, some attributes values may be required in order to get the final decision (4). Once the **PDP** finishes the evaluation, it sends back the response to the **PEP** (5), such response could be *Permit*, *Deny*, *Indeterminate* (if there is an error during the evaluation) or could be *Not Applicable* if there was no applicable policy to evaluate. In the last two cases, the action taken by the **PEP** depends on the type of service to be accessed. Finally, the **PEP** enforces the policy by permitting or denying the access to the resource (6), optionally, the authorization may be accompanied by obligations, which are actions that **PDP** must make on conjunction with enforcing the authorization decision, such as sending an email, log the transaction, or keep the information for a period of time. The scope of the policy definition language is shown as a shade area in the diagram of Figure 3.4.

Policy Enforcement Point (PEP)

As mentioned above, **PEP** is the component that guards access to a set of resources and asks the **PDP** for an authorization decision. If the decision is *Permit*, then the access should be granted, if the decision is *Deny*, the access should be denied, if the decision is *Not Applicable* or *Indeterminate*, then the behavior is undefined. It could be *Permit*, *Deny* or further actions can be taken such as: consultation of additional **PDPs**, reformulation of the decision request, among others.

As mentioned above, **PEP** receives the request and generates a *Decision Request*, which is sent and understood by the **PDP.** The communication protocol and message format between

PEP and **PDP** could be proprietary; however, XACML has already defined a profile in order to bind XACML request/response messages in SAML protocol (Anderson2, 2005).

The *Decision Request* is a collection of attributes that describe a request for access to a particular resource. Basically, the *Decision Request* specifies who is requesting the access over which resource and the actions and purposes under certain environment values. In a general control access model, there are four types of attributes: *Subject* attributes which represent the entity who is making the request, *Resource* attributes that represent the resource to be accessed, *Action* attributes that indicates the operations to be done over the resource and *Environment* attributes that depends on the application context. When this model is applied to a privacy policy framework, an additional type of attribute is required, the *Purpose* attribute which indicates the purpose of the access to the resource. Each attribute has an identifier or name, a data type and a value. Figure 3.5 shows the general structure of a XACML *Decision Request.*

Figure 3.5 Decision Request attributes

An example of Decision Request is shown in Figure 3.6:

Subject Attributes	Resource Attributes	Action Attributes	Purpose Attributes	Environment Attributes
Subject_ID "String" "Researcher"	Resource_ID "URL" "http://int.fr/Project_catalog"	Action_ID "String" "Write"	Purpose_ID String "Update"	Date_ID Date "25/nov/07"

Figure 3.6 Decision Request example

Where **Subject_ID**, **Resource_ID**, **Action_ID**, **Purpose_ID** and **Date_ID** are the names or identifiers of the attributes. **String**, **URL** and **Date** are the data type corresponding to each attribute. **"Researcher", "http://int.fr/Project_catalog", "Write", "Update"** and **"25/nov/07"** are the current values of the attributes.

A single *Decision Request* may contain multiple subjects and resources, depending on the used policy language, therefore, it could be multiple responses, one answer for each resource or only one answer for all the resources. Figure 3.7 shows an example of an *Authorization Decision* corresponding to the *Decision Request* above described.

Resource_ID "URL" "http://int.fr/Project_catalog"	Decision String "Permit"	Obligations String "mailto: owner_ID"

Figure 3.7 Authorization Request example

This response indicates that the action "write" and purpose "Update" to the resource "http://int.fr/Project_catalog" is permitted, but an email must be sent to the owner of the resource indicating the transaction.

Policy Decision Point (PDP)

As mentioned before, **PDP** is the entity responsible to evaluate the policies based on the attributes within the *Decision Request* sent by the **PEP**. In order to carry out its function, there are interactions with additional components as shown in Figure 3.8.

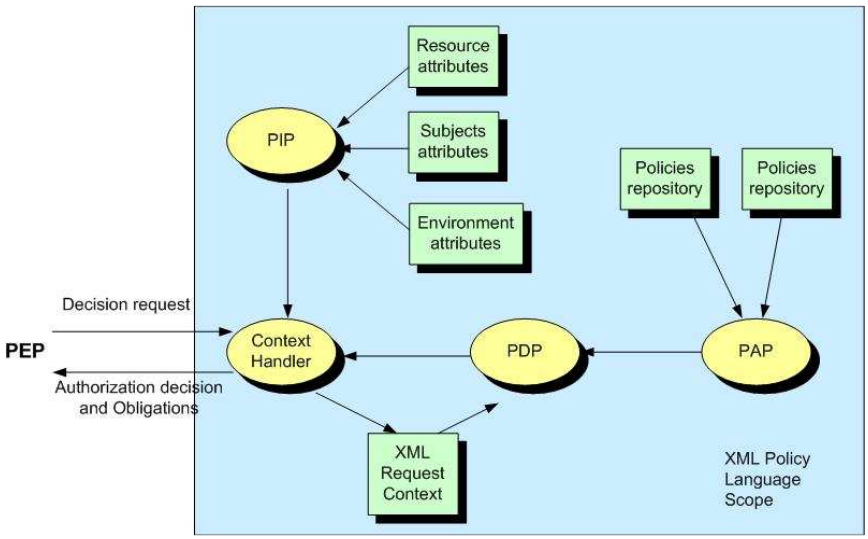


Figure 3.8 Policy Decision Point components

The *Context Handler* is an intermediate component between the **PEP** and **PDP**, mainly because in most cases the **PEP** handles the *Decision Request* in a native format (for example as a character string), so the *Context Handler* translates it to a XML format understood by the **PDP**, additionally, the *Context Handler* could retrieve additional subject, resource and environment attributes as needed during the policy evaluation. The attributes are collected through a component called *Policy Information Point (PIP)* which has direct access to the attributes repositories. *Context Handler* builds an *XML Request Context* with attributes sent by **PEP** as well as local attributes and sends it to **PDP** for the evaluation process. **PDP** selects the applicable policy that matches the attributes within the *Request Context*, then it evaluates the policy and returns a response to the *Context Handler*, which in turn translates the XML response to the native format understood by **PEP**. **PDP** selects the applicable policy through the component called *Policy Administration Point (PAP)*, which manages the policies.

Policy structure

As mentioned above, the PDP matches the Decision Request with the applicable policies in order to evaluate them. The privacy policies have a well defined structure as shown in Figure 3.9. Basically, they are compounded into four sections: the *Target*, the *Combining Algorithm*, the set of *Rules* and optionally the *Obligations*.

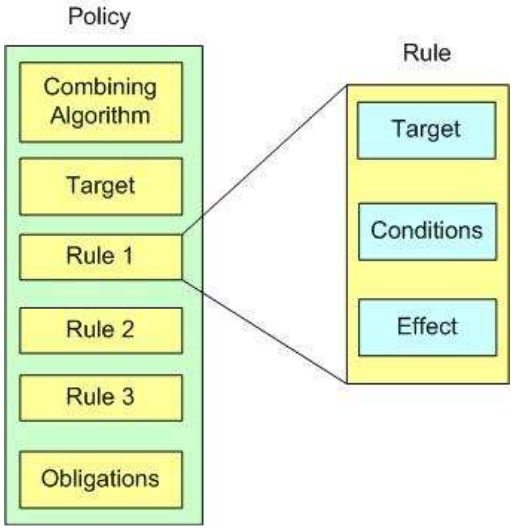


Figure 3.9 Policy and Rule structure

The *Combining Algorithm* determines the way the results of each individual rule are combined in order to get a unique response. The *Target* element defines the applicability of the policy to a particular *Decision Request*, if the subjects, resources, actions, purposes and environment attributes within the *Target* matches the corresponding attributes of the *Request Context*, then the policy may be evaluated by the **PDP** in making its authorization decision. A *Rule* is basically a condition under which access is to be allowed or denied. *Obligations* are optional and specify the actions that must be fulfilled by the **PEP** in conjunction with the authorization decision, such as sending an e-mail or logging the transaction.

Each *Rule* is formed mainly by three elements: the *Target*, *Conditions* and *Effect*. The *Target* in the policy specifies if the policy applies or not. Meanwhile the *Target* within the rule defines if that particular rule is evaluated. *Conditions* are optional Boolean functions over *Subjects*, *Resources*, *Actions*, *Purposes* and *Environment* attributes. If the *Target* of the rule and *Conditions* evaluate to **True**, then the result to be returned by the rule is specified by the *Effect* element, which could be *Permit* or *Deny*. In order to manage complex and distributed privacy policies, some policy languages allow that one policy is made up of multiple sub-policies, which are possibly evaluated and managed separately.

As an example, a University establishes the following privacy policy to protect a web resource:

> **Policy**: "Allow any **Researcher** to **Modify** the *Project Catalog* for keeping **Update** if he or she is the leader of the project. Send an e-mail to the **Researcher** leader when his project registry is modified"

The previous policy described in plain text could be expressed with the following pseudo-code with the structure above explained:

Policy_ID: 1

1.- Combining Algorithm: **Deny_Overrides**
2.- Target:
  2.1.- <Resource>:
    Resource_ID: **"http://int.fr/Project_catalog/*"**
    Data_Type: **"URL"**
  2.2.- <Subject>:
    Group_ID**: "Researcher"**
    Data_Type: **"String"**

2.3.- <Action>:
Action_ID: **"Write"**
Data_Type: **"String"**

Rule_ID: 1

3.- Condition:
String_Not_equal (*Action.Purpose*, "*Update*")
4.- Effect: **Deny**

Rule_ID: 2

5.- Condition:
String_equal (*Researcher_ID*, *Project.Researcher_ID*)
6.- Effect: **Permit**

7.- Obligation:
7.1 mailto("*Project.Researcher_ID*", message)
7.2 message: "Your Project Profile has been accessed by" *Subject_ID*
"for" *Purpose_ID*

The *Target* (2) element of the policy specifies that it applies if the resource named **http://int.fr/Project_catalog/*** of type **URL** (2.1) is requested and also if the requester is a **Researcher** (2.2) that wants to **Write** the resource (2.3). The policy has two rules, the first one controls the purpose of the action, and basically it has to deal with privacy aspects (3), if the purpose is different than "update", the access is denied (4). The second rule verifies that the ID of the **Researcher** is the same in the *Context* and in the *Resource* (5). Finally, if the rule evaluates to *True*, then the value of the *Effect* is returned, in this case the value is *Permit* (6).

If the policy evaluates to *Permit*, then *Obligations* (7) must be sent to **PEP** along with the response. In this case, the obligation is to send an e-mail to the *Researcher* whose project has been accessed (7.1) indicating who is making the request and with what purpose (7.2).

The combining algorithm for this policy is *Deny Overrides* (1) which establishes that if any rule is evaluated to *Deny,* the combined result is *Deny*.

Policy Applicability and Evaluation

For each *Decision Request* sent by the **PEP**, the *Context Handler* creates a *XML Request Context* with the *Subjects*, *Resources*, *Actions*, *Purposes* and *Environment* attributes, such attributes are used by the **PDP** to index the appropriate policy or policies to be applied. In order to retrieve the applicable policies, **PDP** uses a matching function which compares the attributes in the *Request Context* with the attributes specified in the *Target* component of each policy. Those policies whose *Target* matches the *Request Context* are selected as applicable and they are evaluated. The individual results are combined in order to give a single authorization response. A policy with no *Target* element is considered as applicable for all *Decision Request* and then evaluated. If there is no policy that applies, then **PDP** responses with "NotApplicable" *Authorization Decision*. Figure 3.10 shows the flow diagram for a policy matching and evaluation process.

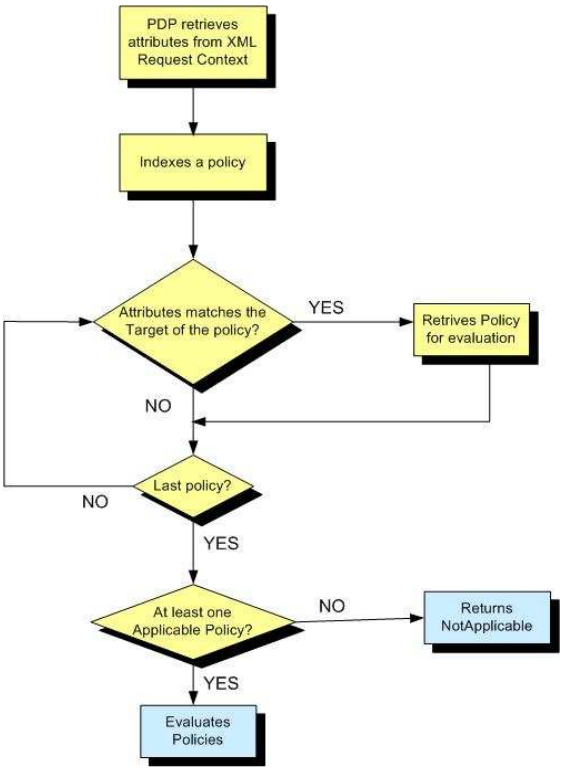


Figure 3.10 Selection of applicable policies

Once a policy is selected as applicable, **PDP** invokes a procedure to evaluate all the rules included within the policy. As mentioned before, a rule is compounded by three main parts: a *Target,* optional *Conditions* and *Effect*. If the *Target* and *Conditions* evaluate to True, then the rule is considered as applicable and the value of *Effect* is returned, either *Deny* or *Permit* response. All the individual results from the evaluated rules are combined as defined in the policy by the *Combining Algorithm* parameter. Finally, a unique response is returned by **PDP** to the *Context Handler* and eventually to the **PEP** entity.

Combining Algorithms

A policy language defines different combining algorithms for arriving at an authorization decision given the individual results of evaluation of a set of rules.

There are a set of standard combining algorithms already defined, which are:

a) **Deny Overrides**.- Establishes that if any rule is evaluated to *Deny,* the combined result is *Deny*.
b) **Permit Overrides**.- Defines that if any rule is evaluated to *Permit*, the combined result is *Permit*.
c) **First Applicable**.- Specifies that the result returned is the value of the first applicable rule, that can be either *Permit* or *Deny*.
d) **Only One Applicable**.- This algorithm applies only to policies, when multiple policies are combined. It ensures that one and only one policy is applicable.

If there are no applicable policies or rules, the result is *NotApplicable*, if there is an error when the policy is evaluated, the result is *Indeterminate*. It is up to the **PEP** implementation to allow or to deny access when a *NotApplicable* or *Indeterminate* answer is returned by the **PDP**.

3.4 Conclusions

There have been many legal initiatives concerning privacy since long time ago, such initiatives have been at local, national or international scope. However, the legal initiative with most impact and which has been used as basis for other works is the 1980 OECD *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.* They propose eight fundamental privacy principles which cover most of the personal information privacy aspects. On the other hand, the increasing number of online transactions and global connectivity through Internet, has incremented the collection and exposure of personal information. Additionally, the use of new technologies (PITs.- Privacy Intrusion Technologies) such as: RFID, biometrics, GPS among others have incremented the risk of privacy for personal information.

Organizations and users are more aware of privacy issues as they are considered as a key element for the development of e-commerce and online collaboration. Technologies can be used in order to support to some extent the privacy principles proposed by the different legal initiatives. Such technologies known as PETs (Privacy Enhancing Technologies) are evolving day after day due to new privacy threats and wide exposure of personal information.

Among all PETs, there is a set of technologies of special interest because they can be applied to information systems with high degree of personal information exposure as the identity management systems, specially the Federated Identity Architecture. Such technologies are the privacy policy languages which allow users and organizations to express, negotiate and enforce privacy policies.

Next chapter will explain how such technologies are used in a general privacy management framework applied to federated architectures.

3.5 Conclusions (en français)

Depuis longtemps, de nombreuses initiatives légales liées au respect de la vie privée ont vu le jour avec une portée locale, nationale ou internationale. Cependant, l'initiative légale la plus marquante et qui a servi de base à d´autres travaux est « 1980 OECD *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* ». Ce guide propose huit principes fondamentaux pour couvrir la plupart des aspects liés au respect de la vie privée des utilisateurs. D´autre part, l´accroissement du nombre de transactions en ligne et la connectivité globale au travers de l´Internet, a accru le risque de collecte et de divulgation d´informations personnelles. En plus, l'utilisation de nouvelles technologies (PITs. – Privacy Intrusion Technologies) comme : RFID, la biométrie, GPS notamment ont augmenté les risques de violation de la vie privée pour les informations personnelles.

Les organisations et les utilisateurs sont de plus en plus sensibilisés à la problématique du respect de la vie privée qui apparaît comme central dans le développement des applications de commerce électronique et des applications collaboratives en ligne. Plusieurs approches issues des différentes initiatives légales sont aujourd'hui disponibles pour mettre en œuvre quelques uns des principes de la vie privée. Ces approches connues sous le nom de PETs (Privacy Enhancing Technologies) sont toujours en cours d'évolution du fait de l'apparition de nouvelles menaces et des expositions importantes des données personnelles.

Parmi tous ces PETs, une approche est toute particulièrement intéressante car elle s'applique à des systèmes d´informations avec un degré élevé d'exposition de données personnelles, comme les systèmes de gestion d´identités, et tout particulièrement les Architectures d´Identité. Il s'agit des langages de politiques de sécurité qui permettent aux utilisateurs et organisations d'exprimer, de négocier et d'appliquer les politiques relatives à la vie privée.

Le chapitre suivant expliquera comment de telles technologies sont utilisées dans un modèle général de gestion de respect de la vie privée appliqué à une architecture fédérée.

Chapter Four

Privacy Model for Federated Identity Management Systems

As described in chapter two, a Federated Identity Architecture (FIA) provides the Single Sign On (SSO) functionality where the user authenticates with the IdP once and can access any service within the CoT with no extra authentication. When the user accesses a service provided by any SP, it is possible that such SP may require additional information about the user in order to improve or personalize the service. The user information requested by the SP, may represent a privacy risk if the data exchanged is sensible personal information (PI). The FIA specifications integrate a set of security mechanisms to guarantee privacy to some extents; however, there are some privacy aspects that are not covered explicitly, so they must be supported by additional functionalities. This chapter explains the inherent privacy mechanisms within the FIA and how they are correlated with privacy principles of a regulatory framework such as the one proposed by the OECD. A layered structured privacy model which is the main contribution of the present work is then proposed in order to complement and enhance the privacy features of the FIA.

4.1 Attribute interchange in a Federated Identity system

Within the Federated Identity system, the exchange of personal information (attributes) can happen between any entity composing the CoT (IdPs, SPs ad users). When the user takes part in the attribute interchange, he can decide to whom release the attribute and under what privacy conditions. However, the most compromised scenario from the point of view of privacy is when the SP requests attributes to the IdP in order to improve and personalize the service. This is because the user has not control over the personal information released by the IdP.

Figure 4.1 shows the possible information flow of the authentication process and attributes interchange between the IdP and SP within a FIA environment. In this scenario the IdP and SP have identity information (ID) of the user, such IDs are federated through a pseudonym. The process starts when the user authenticates with the IdP using any authentication method defined by it (1). If the authentication succeeds, then the IdP gives the user an identity token (2) with authentication information and a pseudonym which is used to access services provided within the CoT. The user requests a service from a SP presenting the token given by the IdP (3), the SP validates the authentication token with the IdP (4) and then maps the pseudo ID to the local identity in order to provide the corresponding service (5). If the service requested needs

additional attributes of the user, they are requested to the IdP (6), the pseudo ID is used to reference the user. The attributes are sent to the SP (7) and finally, the service is granted (8).

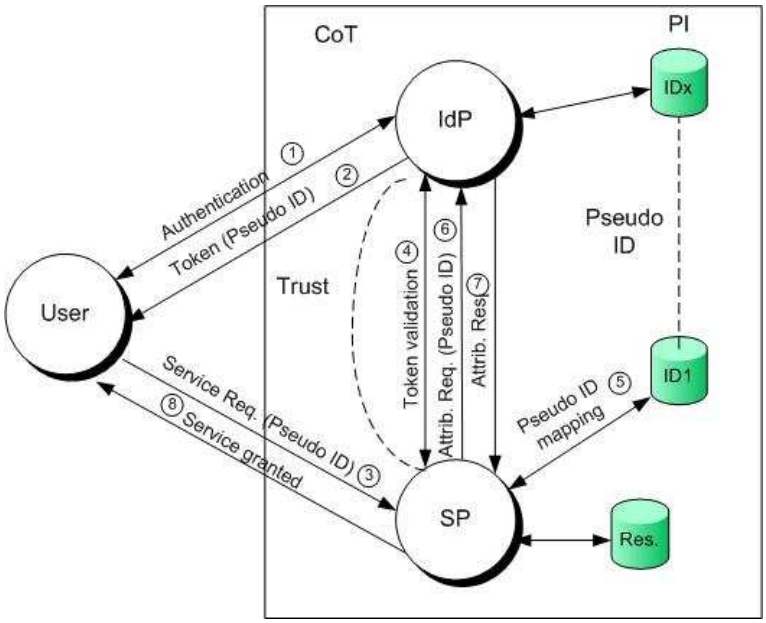


Figure 4.1 A possible attribute interchange in a Federated Identity Architecture

4.1.1 Available privacy mechanisms within the FIA

The FIA provides some tools and guidance to build more secure and private identity services (Varney, 2003), mainly when attributes are transmitted from the IdP to the SP. Such tools and mechanisms are among others:

1. **Channel security**.- The communication between all the components of the FIA is carried out in a secure manner by means of encrypted and authenticated channels, avoiding in this way that any unauthorized entity could intercept personal information.
2. **Message security**.- Additionally to channel security, the messages exchanged by the entities composing a CoT are optionally encrypted and digitally signed, in order to ensure the confidentiality and integrity of the PI exchanged.
3. **Pseudonymity.-** The FIA specification supports the assignment of an arbitrary sequence of characters by the IdP or SP to identify the user (Pseudo ID). It facilitates the identity federation between the user´s accounts at the IdP and SP without knowing the information of the remote account.
4. **Anonymity**.- The FIA specification allows the use of an anonymous ID, which is used to share data to SP for personalized services. This is used when the SP does have a local account for the user and it is not necessary to know his real identity.
5. **Usage Directives**.- The exchanging protocol allows the entities transmitting personal information (*Attribute Requester* and *Attribute Provider*) to integrate directives within the message. These directives can be used to specify the intended use and the allowed usage of the attributes exchanged. The intended use and the allowed usage must match in order to release the corresponding attributes.
6. **Interaction Service**.- FIA includes an *Interaction Service* that enables SPs to make direct interactions with the user to get his explicit consent for certain usage of his attributes.

4.1.2 Missing privacy aspects in Federated Identity systems

The FIA provides the SSO functionality and facilitates the attribute interchange; however, there are many issues regarding how privacy is handled:

- The FIA specifications do not define how personal information (PI) is collected.
- Once PI is collected by any entity, such information is under its control and the user has little or no control on how his PI is released to other entities within the CoT.
- The IdP may optionally notify the user how his PI will be treated in terms of privacy, but there are no enforcing mechanisms to guarantee its fulfillment.
- When the SP requests an attribute, the IdP authenticates the requester and guaranties that the transmitted information is encrypted and digitally signed. However, there are no mechanisms to implement access control.
- Once the PI is released to the SP, the user does not know to whom his personal information is released and for what purposes. There is a lack of an auditing process.

4.1.3 Correlation of privacy mechanisms and privacy principles

The intrinsic privacy mechanisms of the FIA fulfill some privacy principles specified by a regulatory framework; however, there are some missing privacy aspects that must be accomplished with complementary functionalities.

Table 4.1 shows how the privacy mechanisms proposed by the FIA fulfill to some extent the privacy principles proposed by a regulatory framework, in this case, the OECD directives.

	OECD Directives							
	Collection Limitation	Data Quality	Purpose Specification	Use Limitation	Security Safeguards	Openness	Individual Participation	Accountability
FIA intrinsic privacy mechanisms								
Channel Security					■			
Message Security					■			
Pseudonymous				■				
Anonymous				■				
Usage Directives			■			■		
Interaction Services				■				

Table 4.1 Intrinsic privacy mechanisms of FIA and its correlation with the OECD privacy directives

The encrypted and authenticated communications (channel and message security) between the components of the FIA, are two of the most important “Security Safeguards” mechanisms in order to ensure the integrity and confidentiality of personal information interchange. Pseudonymity and anonymity functionalities allow limiting the disclosure of the real identity of the user, fulfilling to some extent the “Use Limitation” principle. Usage directives are exchanged messages to negotiate the intended use of personal information by the requester as

well as the allowed usage of such information by the provider. That functionality can be used to fulfill the "Purpose Specification" and "Openness" privacy principles. Finally, the Interaction Services capability facilitates the user to grant direct disclosure of its personal information, supporting the "Use Limitation" principle. As shown in Table 4.1, there are some privacy principles that are not fulfilled by the intrinsic privacy mechanisms of FIA.

The following section proposes a Privacy Management Model based on privacy policies. The model considers new components and functionalities that must be integrated with the entities of the FIA responsible for sharing personal information (the Attribute Provider). The model satisfies the privacy principles that are not fulfilled by the intrinsic FIA mechanisms.

4.2 Privacy Management Model

The proposed privacy model has as main objective to complement and to reinforce the privacy aspects of the FIA in order to technically support the privacy principles of a legal framework such as the OECD directives. The basic idea is to protect the PI by means of privacy policies taking into account regulatory requirements (external and internal) and the user privacy preferences. The model must enforce the privacy policies and must allow the auditing of its main functionalities which are necessary to improve the level of privacy within a FIA system; the model provides a set of functionalities that allow:

1. To define privacy requirements
2. To represent personal information in a standard data abstraction format.
3. To access personal information for verifying and updating purposes.
4. To create privacy policies.
5. To specify privacy preferences of users.
6. To enforce privacy policies.
7. To audit the compliance of privacy policies.

Privacy policies are defined in a hierarchical way taking into account national and international regulations, organizational policies and user preferences. The high level privacy definitions must be mapped from natural language expressions to policy using privacy expression languages so that entities exchanging personal information can handle the same syntax and semantics of policies. Privacy policies are defined at the entity side providing the personal information (Attribute Provider, AP) and optionally at the entity side requesting the attributes (Attribute Requester, AR). Each activity at the privacy policy definition and enforcement must be logged so that an audit system could verify them. This functionality is very important for the users to know at any time how their personal information is handled; similarly, the organizations could prove compliance with the corresponding privacy regulations.

Before explaining the architecture and functionality of the model, there are initial assumptions that we must consider; the organizations exchanging personal information belong to the same CoT and they have established a business agreement regarding privacy policy mechanisms and a compatible regulatory framework. The following sections explain the general architecture, the functionality of the components and the relationship between them.

The architecture proposed is a 3-layer model shown in Figure 4.2. The lower layer, named *Privacy and Personal Information Metadata* layer, is responsible for privacy requirements specifications and the PI definition and handling (all the specifications at this level are done using metadata and ontology expression languages). An entity known as the Privacy Officer (PO) provides the privacy specifications in terms of external and internal privacy requirements. This layer is also the interface to multiple PI sources. The intermediate layer, known as *Policy Services* layer, is in charge of creating, handling and enforcing the privacy policies. This layer allows the users to specify their privacy preferences and to verify the policies associated with their PI. The upper layer or *Attribute Services* layer responds to *Attribute Requesters* and to users requesting access to personal information. All the activity that happens at the two lower layers is logged by a module named *Personal Information and Policy Log*. This information is

available to the PO for auditing purposes and to the user so he could know at any time how his PI is handled by the AP in terms of privacy.

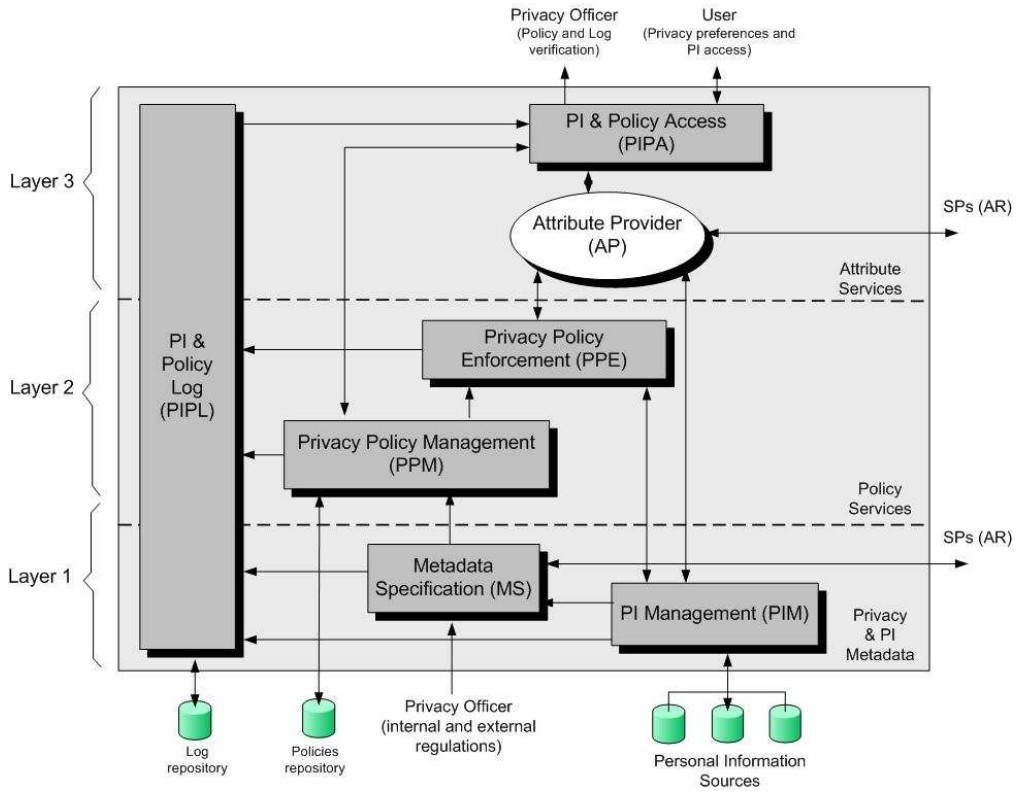
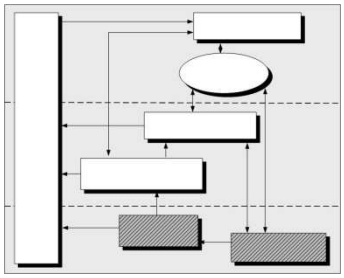


Figure 4.2 Privacy model architecture

The following sections will describe with more detail the components and functionality of each layer.

4.2.1 Privacy and Personal Information Metadata layer



This layer provides basic functionalities that allow to specify privacy requirements and to handle personal information. To accomplish these functionalities, the layer is composed of two modules: The *Metadata Specification* (MS) module and the *Personal Information Management* (PIM) module. The MS module receives privacy requirements from the privacy officer and generates a privacy ontology that correlates personal information, services and privacy requirements. The PIM module provides a standardized interface to heterogeneous personal data sources. The following subsections describe the architecture and functionality of each module.

Metadata Specification module

This module is composed of two Web Services: the *Ontology Merging Web Service* (OM-WS) and the *Ontology Exchange Web Service* (OE-WS).

The OM-WS creates privacy profiles (structured in a *Privacy Policy Ontology* or PPO) which correlates personal information profiles (defined within a *Personal Information Profile Ontology* or PIPO) and service profiles (specified in a *Service Profile Ontology* or SPO) with privacy requirements. The PI profiles are created by the *Personal Information Management* module (explained later) of the same layer, the service profile definition is generated by each SP

with whom the *Attribute Provider* will exchange personal information; the privacy requirements are input in a natural language format by the *Privacy Officer*.

The OE-WS exchanges ontologies with its peer at the *Attribute Requester* side. This process is carried out at the initial interaction between the AP and the AR or each time any ontology definition changes. The PPO is sent to the *Privacy Policy Management* module since it works as a baseline for privacy policy generation. Each time a new PPO is created or modified, the action is logged.

The *Metadata Specification* module supports some privacy principles of the OECD such as the "Purpose Specification" due to the association of privacy requirements to personal information profiles. It also supports the "Accountability" principle because the logging of metadata creation facilitates the auditing process. Figure 4.3 shows the web services and the data exchanged within the MS module.

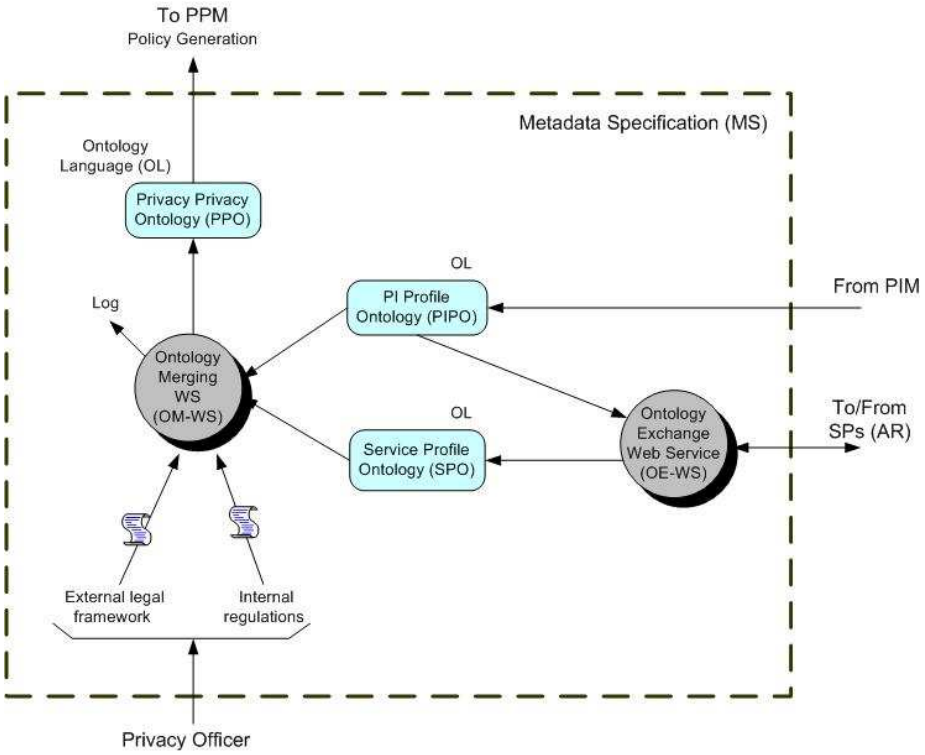


Figure 4.3 Metadata Specification module

As shown in Figure 4.4, the PI Profile Ontology generated by the PIM module, contains PI profiles with categorized information (a category is a set of related personal information). The Service Profile Ontology, which is sent by the SP, describes service profiles offered by the SP where each profile is formed by a set of individual services. The Privacy Officer associates PI profiles with Service profiles assigning different privacy requirements to each of the personal information category, generating in this way a Privacy Policy Ontology which is the baseline for the privacy policies construction.

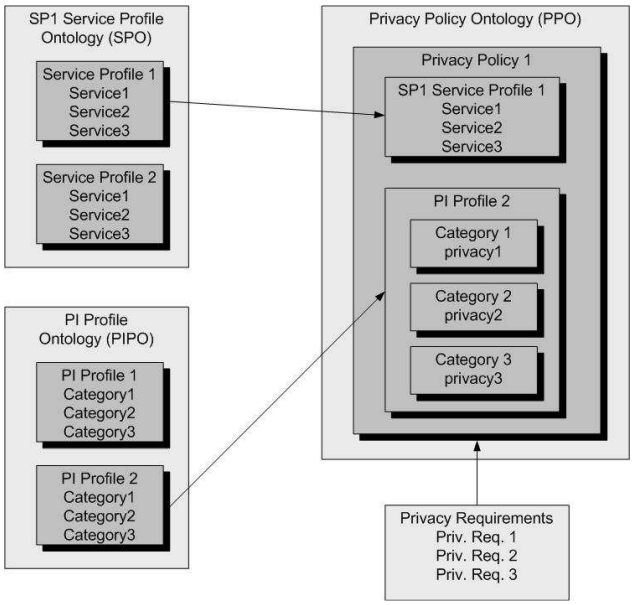


Figure 4.4 Privacy Policy Ontology

It is important to have a good semantic understanding of personal information, service description and privacy requirements in order to guarantee a consistent privacy policy building. All ontologies are written with the same ontology language initially negotiated by OE-WS between the peers.

Figure 4.5 shows the interactions between the web services (shaded components) of the *Metadata Specification* module and external entities during the privacy profile construction.

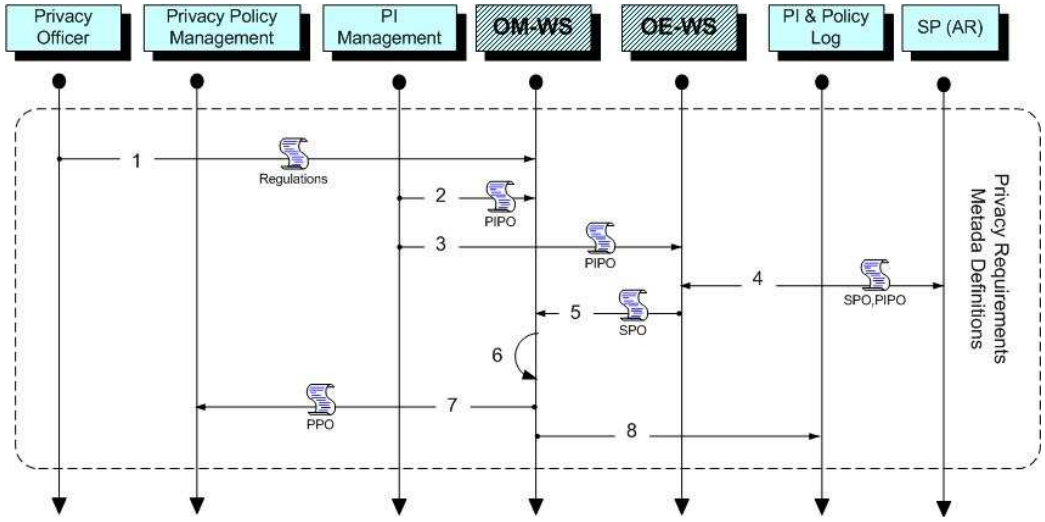


Figure 4.5 Metadata specification process

The *Privacy Officer* specifies the external and internal privacy regulations (if they exist) in natural language format (1). The *Personal Information Management* module (explained in the next section) generates and sends the PI profiles to the OM-WS and OE-WS for privacy ontology creation and ontology interchanging respectively (2 and 3). The OE-WS exchanges ontologies with its peer at the *Attribute Requester* side (4), sending the PIPO and receiving the SPO. The SPO is sent to the OM-WS (5) for ontology merging. The OM-WS combines all the ontologies with the privacy regulations (6), creates the *Privacy Policy Ontology* (PPO) and sends it to the *Privacy Policy Management* module (7) for policy creation. The creation of PPO is logged into the *Personal Information and Policy Log* module.

The metadata specification process is carried out once before any attribute exchange between the AP and AR, or each time a change occurs in the privacy regulations, PI or service profiles.

Personal Information Management module

The *Personal Information Management* (PIM) module works as an interface with multiple and heterogeneous personal information sources. It is composed of two web services; the *Data Mapping Web Service* (DM-WS) and the *Personal Information Web Service* (PI-WS).

There is a DM-WS for each data source generating a standard data abstraction known as *Personal Information Data Abstraction* (PIDA). This mapping functionality is important to homogenize the syntax and semantics of PI and to provide a transparent and consistent data access from any entity requesting personal information. Secure functionalities could be added to this web service such as data encryption and digital signature as safeguards mechanisms for the stored data.

PI-WS is the interface between the homogenized personal information and the modules requesting attributes. The *Attribute Provider* and the *Privacy Policy Enforcement* modules request PI without knowing the source or the format of the original data. Additionally, this web service takes the data abstraction generated by each DM-WS and builds a *Personal Information Profile Ontology* that describes structured personal information profiles represented by data categories. Such categorization facilitates the correlation of PI with services and privacy levels. The PI ontology is sent to the *Metadata Specification* module for privacy profile construction. Each time PI is accessed, the operation is logged so the user could know at any time how his PI is handled.

The PIM module gives support to the following privacy principles: “Data Quality” and “Security Safeguards” by guarantying the integrity and consistency of personal information, “Individual Participation” by allowing the user to access his own personal information, and “Accountability” by logging all the accesses to PI so the users could know how their personal information is used. Figure 4.6 shows the web services and data exchanged within the PIM module.

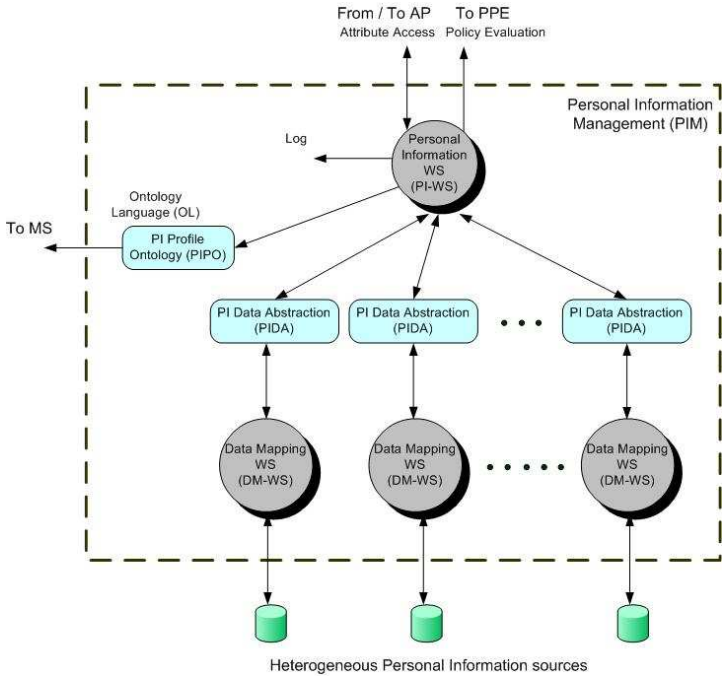


Figure 4.6 Personal Information Management module

Figure 4.7 shows the interactions between the web services (shaded components) of the PIM module and external entities during the process of PI data abstraction and PI access.

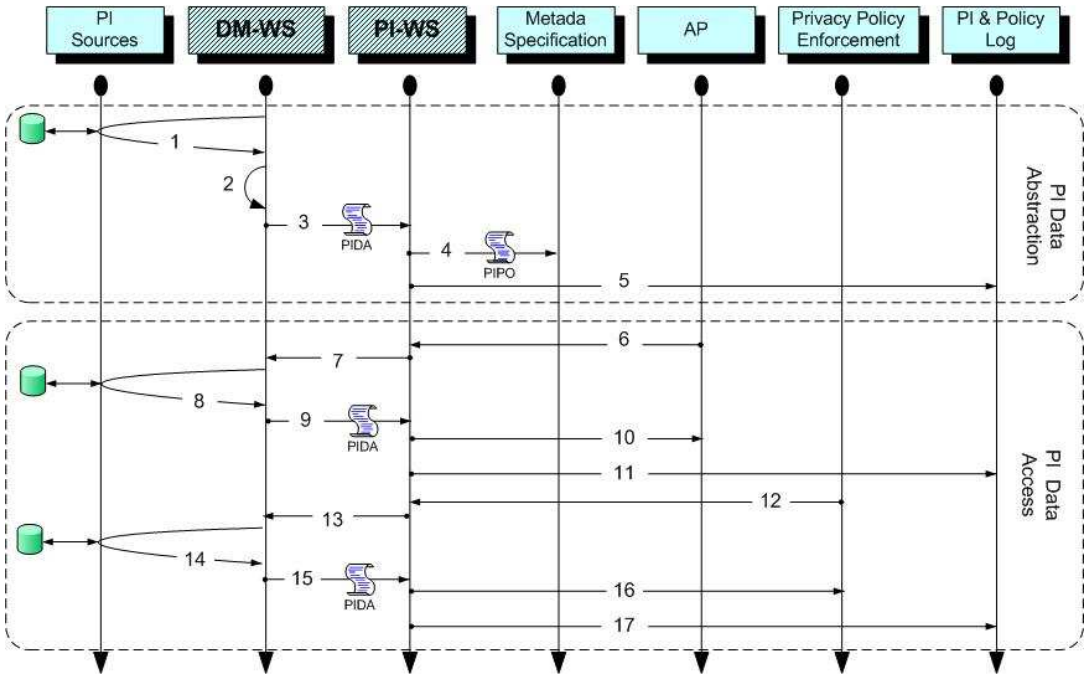
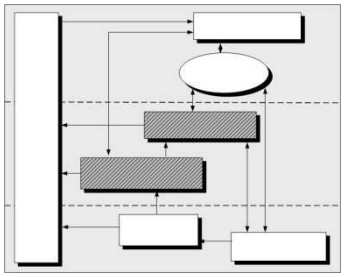


Figure 4.7 PI abstraction and Access processes

For the PI data abstraction process, the DM-WS connects to heterogeneous data sources (1) and proceeds to map the original data to a common data abstraction (2), guaranteeing the syntactic consistency of the information. The PI data abstraction is read by the PI-WS (3) in order to generate the *Personal Information Profile Ontology* which is sent to the *Metadata Specification* module (4). Each time a data source structure is modified or a new source is added, the PI data abstraction and PI profile ontology are modified.

The data access process starts with the personal information request from the *Attribute Request* or the *Privacy Policy Enforcement* modules (6,12) to the PI-WS, which turns out the request to the DM-WS (7,13) for direct access to the data sources. The personal information is retrieved from the corresponding data source (8,14), mapped and sent back to PI-WS in a data abstraction format (9,15). Finally, the PI is sent to the original requester (10,16). All activity of the PI-WS during data abstraction and PI access process is logged for auditing purposes (5,16 and 17).

4.2.2 Policy Services layer



The *Policy Services* layer deals with the creation and handling of privacy policies. It contains two modules: The *Privacy Policy Management* (PPM) module and the *Privacy Policy Enforcement* (PPE) module. The PPM module creates the privacy policies allowing the users to specify their privacy preferences. It also indexes, stores and retrieves policies as they are requested by other modules. The PPE module intercepts the attribute request; it matches the request to the corresponding privacy policy and then proceeds to evaluate it. The result of the evaluation is sent back to the AP entity in order to grant or deny the attribute. The following subsections describe the architecture and functionality of each module.

Privacy Policy Management module

This module is responsible for the creation and handling of privacy policies, it contains two web services: The *Policy Generator Web Service* (PG-WS) and *Policy Storing and Indexing Web Service* (PSI-WS).

The PG-WS receives the *Privacy Privacy Ontology* from the *Metadata Specification* module and generates an initial *Privacy Policy Set* (PPS). The PPS is defined in a standard *Privacy Expression Language* (PEL). This policy set is associated with each personal profile defined by the organization; it is built taking into account the regulatory privacy requirements as well as the privacy preferences from the users expressed in a standard *Privacy Preference Language* (PPL). If the user does not specify any privacy preference, a default preference is assigned by the *Privacy Officer*. The policy is evaluated and enforced each time an attribute is requested. The PPS is available at any time so users and privacy officers can know the privacy policies associated to each personal profile. The PPS is sent to the PSI-WS for storing and indexing purposes. Each time a PPS is generated or modified the action is logged.

The PSI-WS stores and indexes privacy policies during the process of policy construction. It also retrieves privacy policies requested by the *Privacy Policy Enforcement* module and interacts with the *PI and Policy Access module* so that users can know at any time the privacy policies associated with their personal information. All the policy handling actions are logged for auditing purposes.

The PPM module supports the following privacy principles: "Openness" and "Purpose Specification" because the organization lets the users know the institutional privacy policies and how personal information will be handled in terms of privacy, "Use Limitation" since the user can express his privacy preferences, and the "Accountability" principle because the user and privacy officer could know at any time the policies associated to the PI profiles. Figure 4.8 shows the web services and data exchanged within the PPM module.

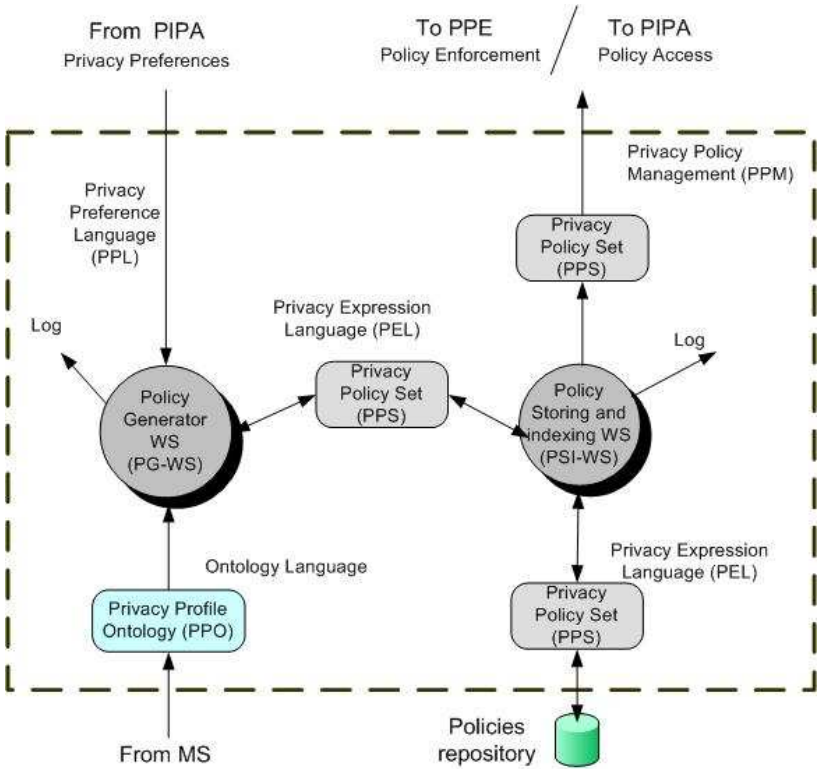


Figure 4.8 Privacy Policy Management module

Figure 4.9 shows the interactions between the components of the PPM module during the processes of privacy preference specification and policy access.

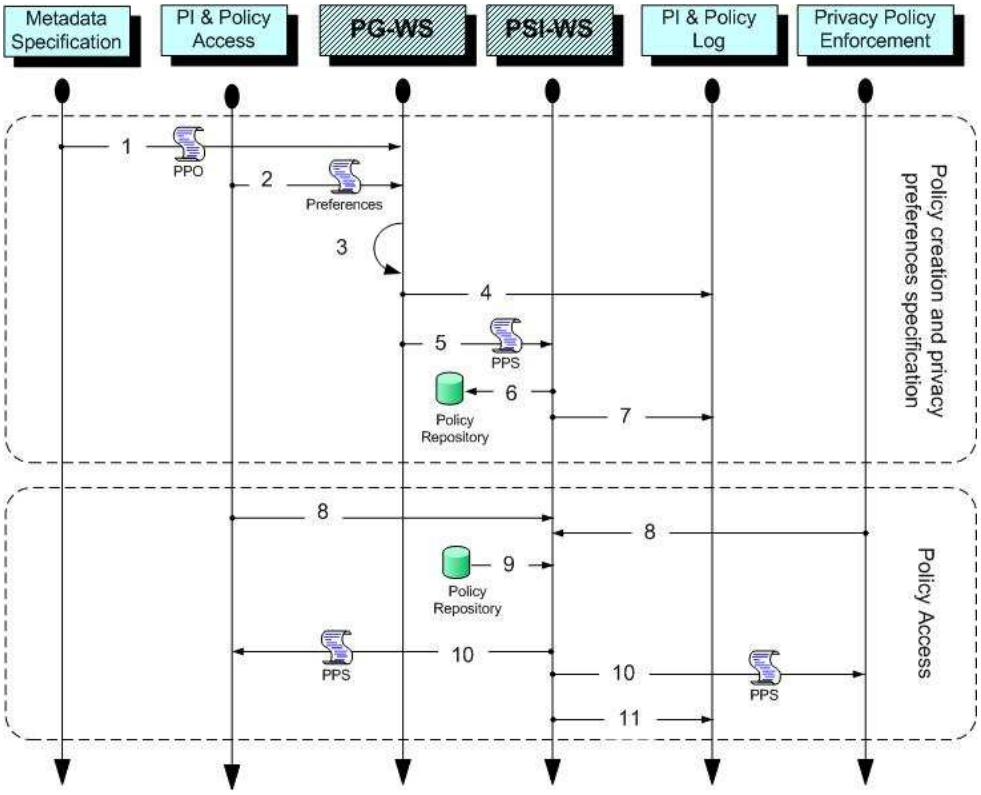


Figure 4.9 Policy creation, preference specification and policy access processes

The PG-WS receives the *Privacy Policy Ontology* (1) and optionally the privacy preferences from users (2) and creates the *Privacy Policy Set* (3). The PPS is sent to the PSI-WS (5) for storing and indexing purposes (6). Each time a new privacy policy is created, modified or stored, the operation is logged (4,7). For the policy access process, the policy request could be from the policy interface (*Privacy Officer* or user) or from the *Privacy Policy Enforcement* module (8). The policy is retrieved (9) and sent to the policy requester (10). Each time a policy is requested, the action is logged (11).

Privacy Policy Enforcement module

The *Privacy Policy Enforcement* (PPE) module is responsible for enforcing the privacy policies at the *Attribute Provider* side. It is composed of two web services: the *Policy Enforcement Point Web Service* (PEP-WS) and the *Policy Decision Point Web Service* (PDP-WS).

The PEP-WS is the web service that intercepts the request generated by the *Attribute Requester* of the SP which is redirected by the AP. It sends this request to the PDP-WS for policy evaluation. The resulting decision is notified to the *Attribute Provider* in order to grant or deny the request. Additional obligations could be sent to the *Attribute Requester*; an obligation is a directive that specifies how the attribute must be treated by the AR in terms of privacy.

The PDP-WS receives the attribute request. It analyses the request and finds the *Privacy Policy Set* that matches the parameters of the request. Such policy set is retrieved from the *Privacy Policy Management* module. If additional attributes are required for the evaluation process, they are requested to the *Personal Information Management* module. The privacy policy is evaluated and the result is notified to the PEP-WS. All policy evaluation and enforcement are logged for auditing purposes.

The PPE module supports the following privacy principles: “Security Safeguards” and “Use Limitation” because it implements access control mechanism to PI and enforces the compliance

of privacy policies. “Accountability” principle is also supported due to the log of each action. Figure 4.10 shows the web services and data exchanged within the PPE module.

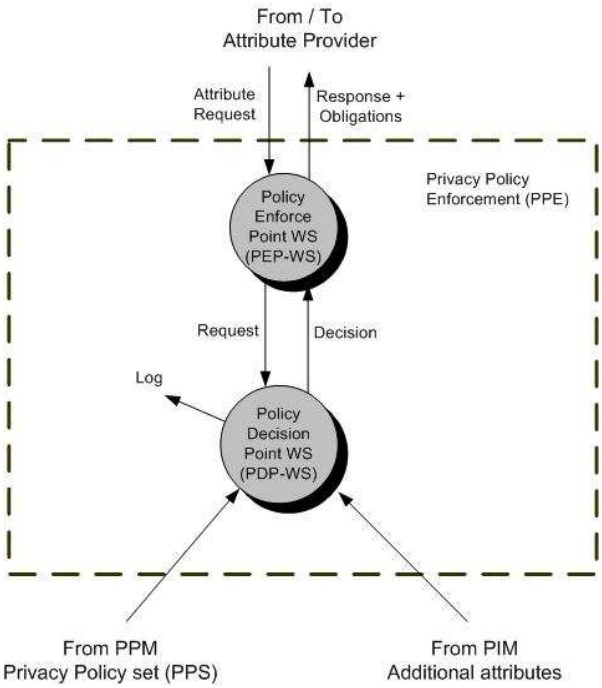


Figure 4.10 Privacy Policy Enforcement module

Figure 4.11 shows the interactions between the components of the PPE module during the process of policy evaluation and enforcement.

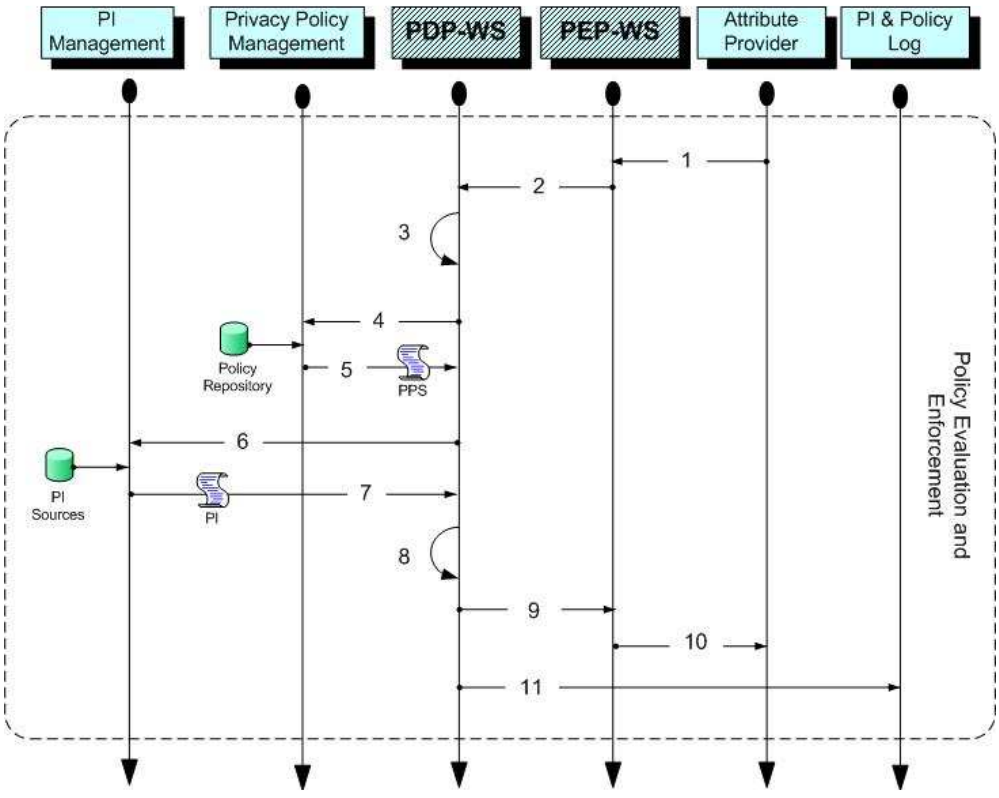
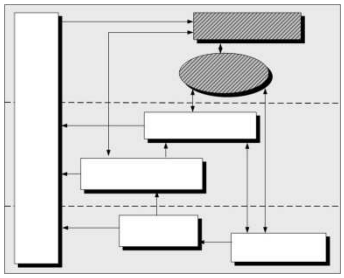


Figure 4.11 Policy evaluation and enforcement processes

The *Attribute Provider* redirects the personal information request coming from the *Attribute Requester* to the PEP-WS (1); this request is converted to a standardized representation and sent to the PDP-WS for evaluation (2). The PDP-WS analyses the request (3) and looks for any privacy policy that matches the parameters of the request (4). The privacy policy is retrieved from the policy repository by the PPM module and sent back to the PDP-WS (5). If additional personal information attributes are required for the evaluation process, they are requested to the PIM module (6,7). With all the necessary information, the PDP-WS proceeds to evaluate the privacy policy (8). The result is sent to the PEP-WS with possible obligations regarding how the attributes must be treated if the response was affirmative, or the reasons if the response was negative (9). The PEP-WS sends the response to the *Attribute Provider* for granting or denying the attributes (10). All the actions taken by the PDP-WS are logged for auditing purposes (11).

4.2.3 Attribute Services layer



The *Attribute Services* layer provides access to personal information requested by the *Attribute Requester* of a *Service Provider,* or directly by the user and the *Privacy Officer*. This layer is composed of the *Attribute Provider* (AP) entity and the *Personal Information and Policy Access* (PIPA) module. The AP entity is an original component of the IdP and it is responsible for releasing the attributes requested by an entity requester.

The PIPA is an interface that allows the user to access his personal information in order to update it or to verify its consistency and integrity. The PIPA supports the following privacy principles: "Collection Limitation" because the user can control the amount of personal information collected by the AP of the IdP, "Data Quality" since the user is able to verify that his PI is accurate, complete and up to date, and "Individual Participation" since the user can access directly its own personal information. The interface also allows the user and *Privacy Officer* to verify the privacy policies associated to the personal information, supporting in this way the "Openness" principle. Finally, a log system can be consulted, therefore, the user and *Privacy Officer* can know at any time how the personal information is handled, supporting the "Accountability" privacy principle.

Attribute Provider

An *Attribute Provider* is an entity that provides attributes to a requester (i.e., a Service Provider) in accordance with its own policies and user's permissions. Attribute Providers store and negotiate access control information defining the circumstances under which a Service Provider will be granted access to a given attribute. *Attribute Providers* store and negotiate usage directives that specify the manner in which attributes can be used, stored, and disclosed. An *Attribute Provider* has at least the same responsibilities as *Service Providers* with respect to clear notice (including notice to the user regarding what are the default usage directives and how the user can change such usage directives), choice, security, and responsible use and sharing of user's data (Varney, 2003).

Within the privacy model proposed, the AP does not deliver the PI on its own. It redirects the attribute request to the *Privacy Policy Enforcement* module.

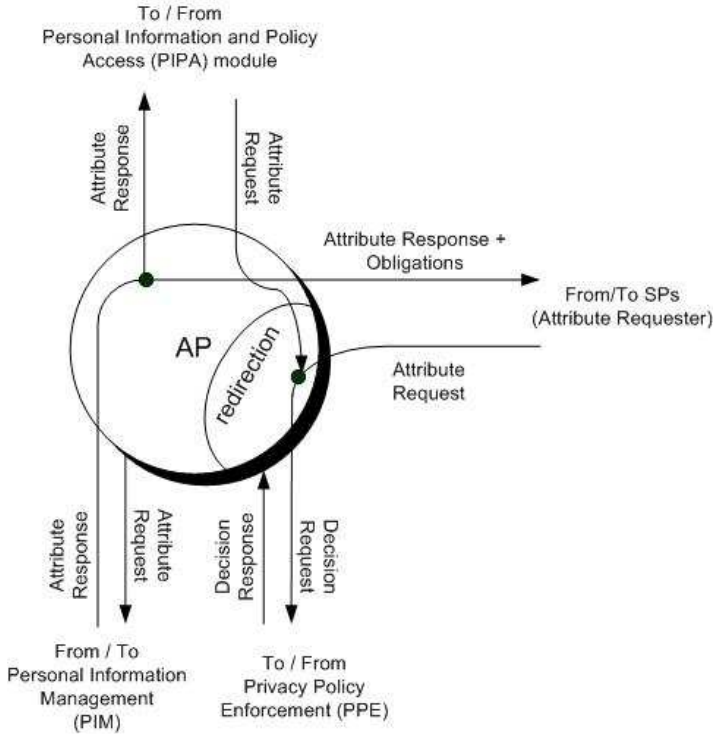


Figure 4.12 Attribute Provider functionality

As shown in Figure 4.12, the AP could receive attribute requests from SPs or directly from users through the *Personal Information and Policy Access* module. In any case, the AP redirects the requests to the PPE module for policy enforcement purposes. This assures that privacy policies are enforced for all the requested attributes. The request could be either consult or modification of the attributes. If the policy evaluation results are positive, then the attribute is retrieved from the *Personal Information Management* module and sent to the SP or the PIPA module.

Figure 4.13 shows the interactions between the Attribute Provider and other entities when handling the attribute request and attribute response.

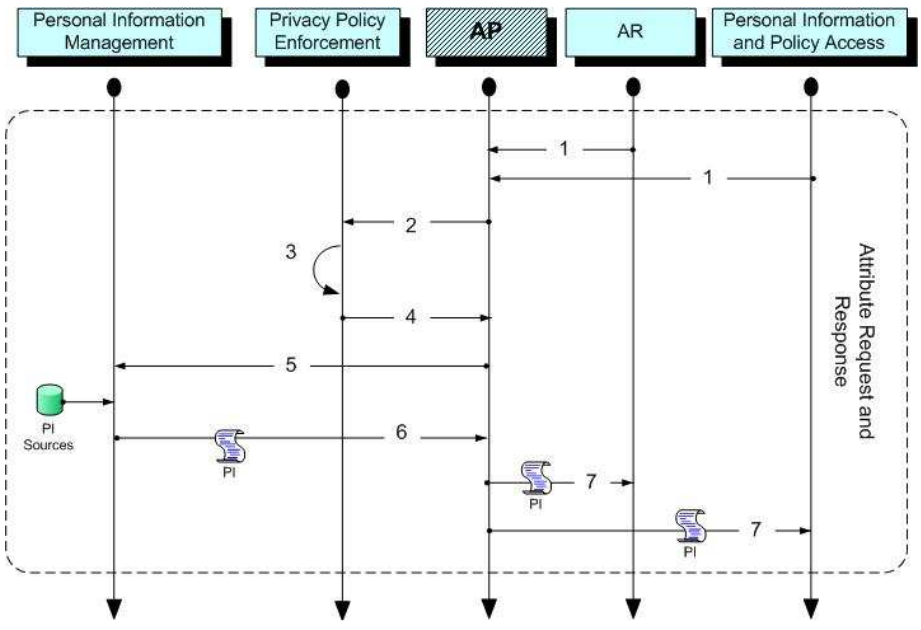


Figure 4.13 Attribute request and response handling

The AP receives the attribute request from the AR entity of the corresponding SP or from the *Personal Information and Policy Access* module (1). The request is redirected to the *Privacy Policy Enforcement* for policy evaluation (2); the request is evaluated (3) and the result is sent back to the AP (4). If the result is positive, the AP requests the attribute from the *Personal Information Management* module (5); the attribute is retrieved from the PI source and sent back to the AP (6), which in turn transmits it to the requester entity or module (7).

Personal Information and Policy Access module

The Personal Information and Policy Access (PIPA) module allows users and Privacy Officers to access personal information, privacy policies and event logs. The PIPA module translates the internal data structures (i.e. policies, attributes and logs) to a human readable format. The PIPA module is composed of a *Graphical Interface* (GI) and three web services: *Log Interface Web Service* (LI-WS), *Data Interface Web Service* (DI-WS) and *Policy Interface Web Service* (POI-WS).

Through the GI, the user can access his personal information and express his privacy preferences. Additionally, privacy officers and users can consult the privacy policies associated to the personal information and the event logs generated by the functionality of the entire model. The LI-WS retrieves the event logs requested by users or privacy officers and presents them in a human readable format; classified in log classes. The DI-WS requests attributes to the AP, so that users can access their personal information during the collecting, consulting and updating processes. POI-WS allows users to express their privacy preferences in a natural language format and transforms them into expressions in privacy policy language. Similarly, it retrieves the policies specified in privacy expression language and transforms them into natural language for easy understanding to users and privacy officers.

Figure 4.14 shows the graphical interface, web services and data exchanged within the PIPA module.

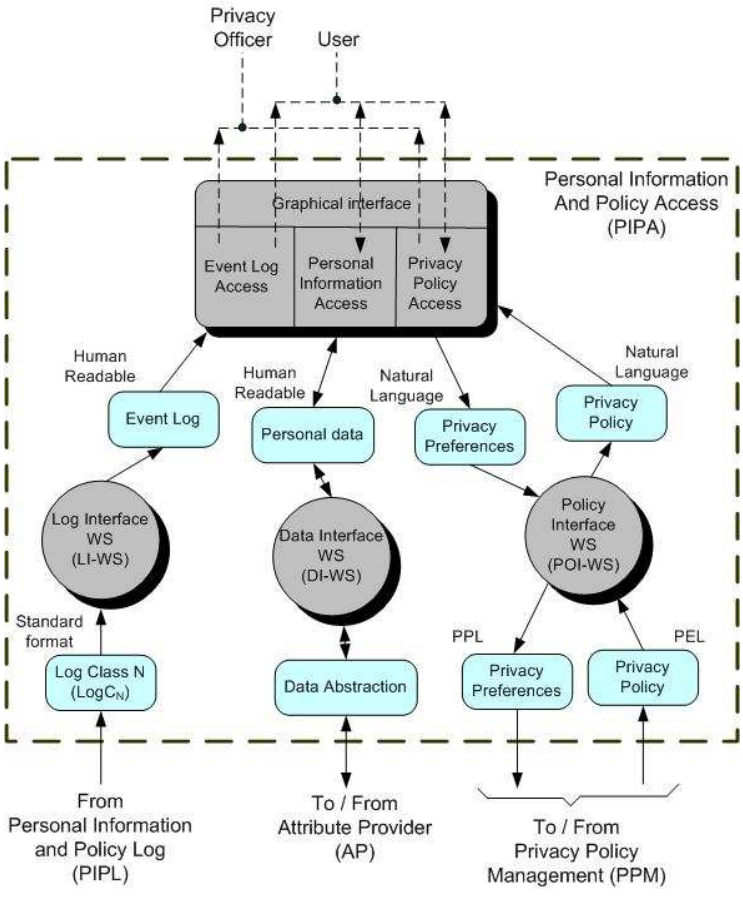


Figure 4.14 Personal Information and Policy Access module

Figure 4.15,Figure **4.16** andFigure **4.17** show the interactions between the components of the PIPA module when handling the event log, personal information and privacy policy access respectively.

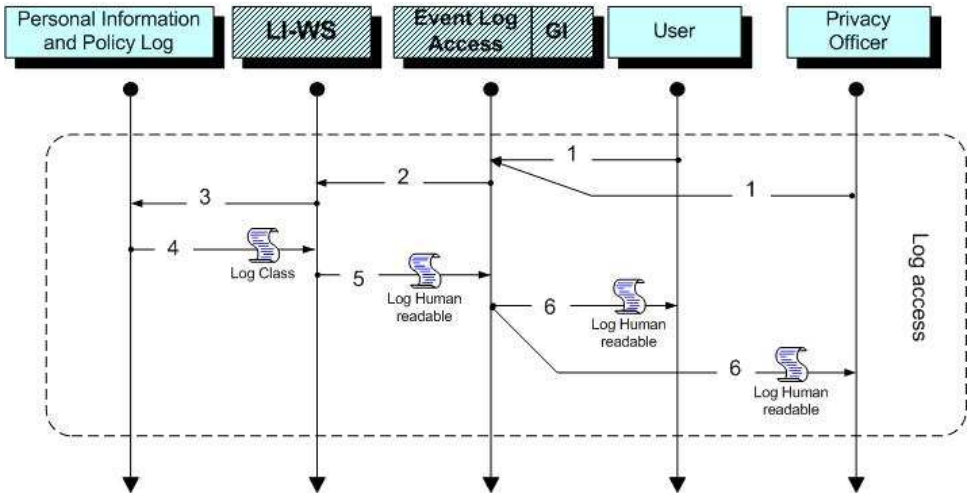


Figure 4.15 Event log access handling

Users and privacy officers use the *Event Log Access* sub-module of the graphical interface to select the type of event log (log class) to be consulted (1). The selection is sent to the LI-WS (2), which maps the selection to a request towards the *Personal Information and Policy Log* module (3). The log class is retrieved and sent back to the LI-WS (4). The log class expressed in an internal format is translated to a human readable format (5) and presented to the users or privacy officers through the GI.

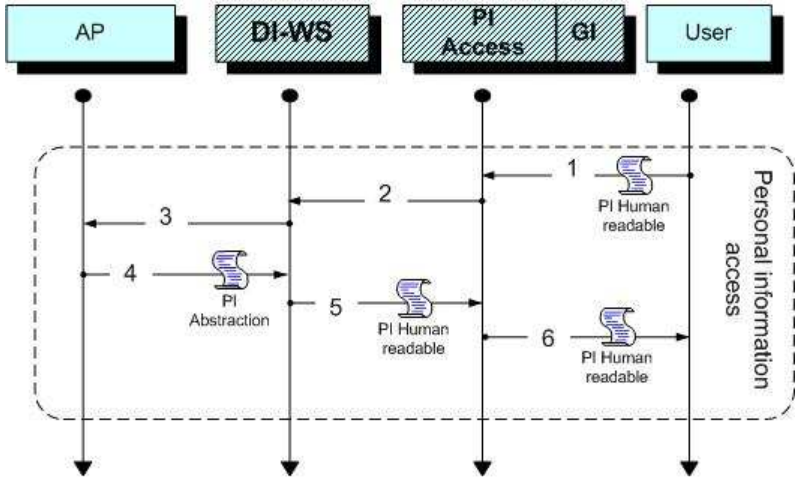


Figure 4.16 Personal information access

Figure 4.16 shows the data flow when a user accesses the *Personal Information and Policy Access* module in order to: enter, consult or modify his personal information. The user selects the PI category to be accessed using the *PI Access* sub-module of the graphical interface (1); the selection is passed to the DI-WS (2) so it can build and send the request to the AP (3). The AP retrieves the PI expressed in a data abstraction format (4); the DI-WS converts the PI to a human readable format (5) for presenting it to the user via the GI (6). This operation, allows the user to limit the amount of PI to be collected by the IdP, in addition, he can verify its consistency and integrity.

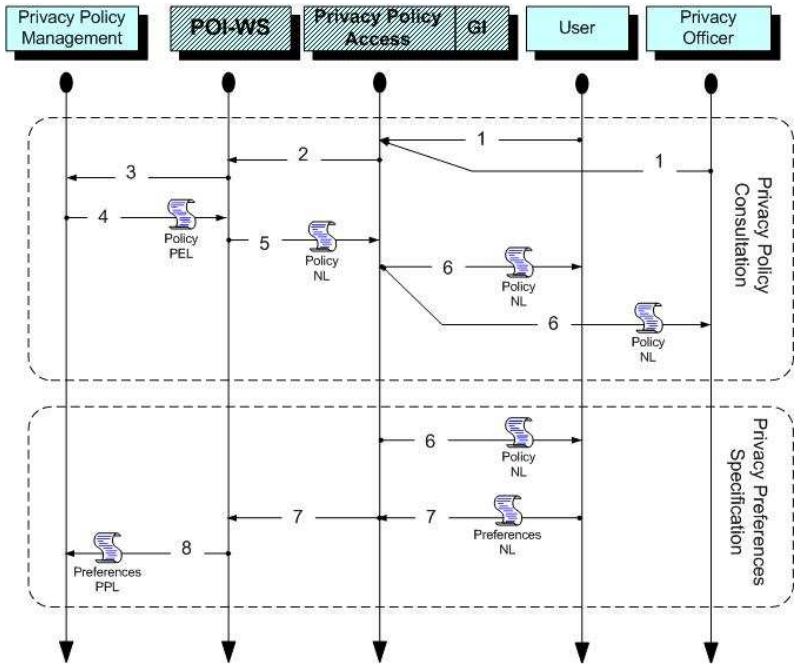
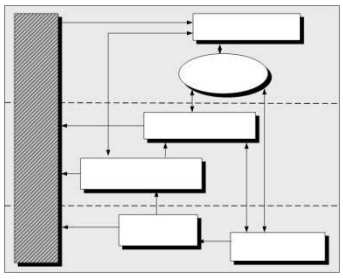


Figure 4.17 Privacy policy consultation and preferences specification

Privacy policies can be consulted by users and privacy officers using the *Privacy Policy Access* sub-module of the graphical interface of the PIPA module (1). The privacy policy selected is indicated to the POI-WS so it can generate the corresponding policy request to the *Privacy Policy Management* module (3). The privacy policy expressed in any privacy expression language (Policy PEL) is retrieved by the PPM module and sent back to the POI-WS (4). The POI-WS translates the privacy policy to a natural language format (5) so that users and privacy officers can understand it (6). This operation allows them to verify at any time which privacy policy is associated to the personal information.

Once the user accesses the privacy policy (6), he can optionally modify it by specifying his privacy preferences only for certain personal information categories (7). Finally, the POI-WS converts the preferences expressed in the natural language to privacy preference language (8).

4.2.4 Personal Information and Policy Log module



The Personal Information and Policy Log (PIPL) module interacts with components of the entire architecture at all layers. The PPL module is composed of two Web Services: the *Log Classification Web Service* (LC-WS) and the *Log Store Web Service* (LS-WS).

The LC-WS receives event logs from *Privacy and PI Metadata* layer when ontologies are created or modified and when personal information is retrieved from the PI sources. Similarly, it receives event logs from the *Policy Services* layer when privacy policies are created or modified and when policies are enforced. Event logs are classified (into Log Classes) and formatted (using any internal format) for storing purposes. Users and privacy officers can verify at any time how personal information is handled in terms of privacy; additionally, an automated auditing process[4] can be implemented for assuring regulatory compliance.

The LS-WS receives the classified and formatted event logs from the LC-WS and proceeds to index and store them in a local log repository.

[4] Out of the scope from the present work

Figure 4.18 shows the web services and data exchanged within the PIPL module and their communication with the components of the three layers.

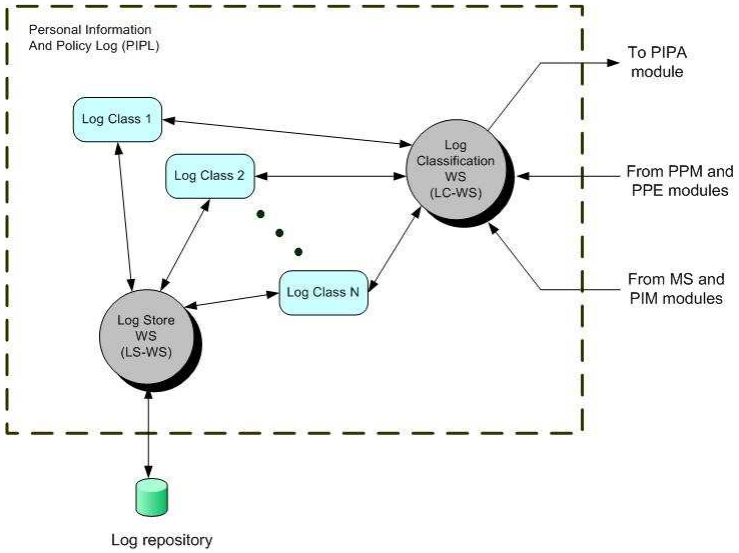


Figure 4.18 Personal Information and Policy Log module

Figure 4.19 shows the interactions between the components of the PIPL module during the log request and log store operations.

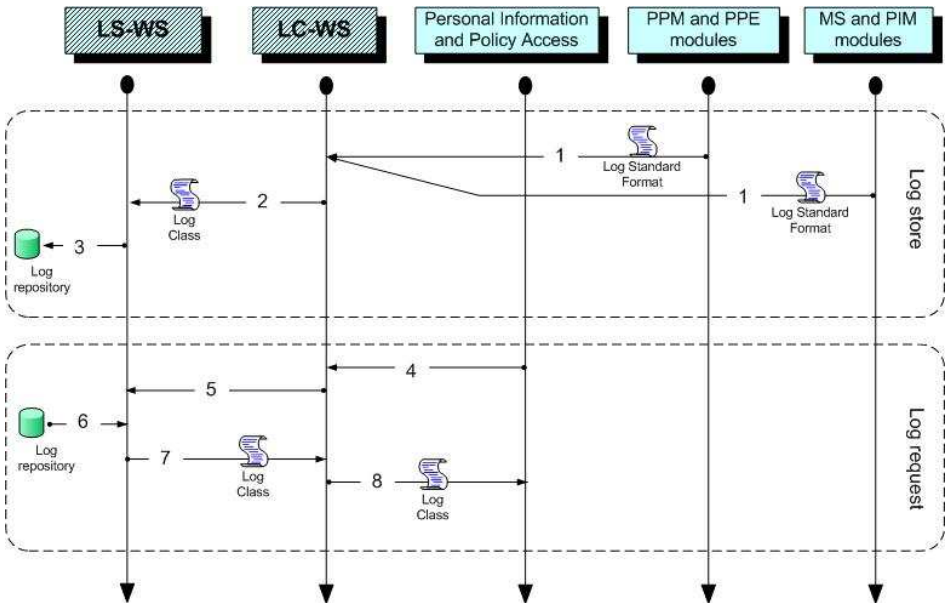


Figure 4.19 Log store and log request operations

The LC-WS receives the event logs from the modules of the lowest and intermediate layers (1); it classifies the event log into the corresponding log class and sends it to the LS-WS for storing purposes (2). Finally the event log is stored within the log repository (3).

For log consulting, the LC-WS receives the event log request from the *Personal Information and Policy Access* module (4); it classifies the corresponding log class and sends the request to LS-WS (5). The LS-WS retrieves the event log from the log repository (6) and sends it back to LC-WS (7) which in turn delivers the log class to the PIPA module (8).

If an automated auditing system is implemented, the request for event log consulting can come from other entities besides the PIPA module.

4.3 Privacy Management Model and the privacy principles

The privacy model previously presented complements the intrinsic characteristics of the FIA in order to enhance the privacy compliance of a regulatory framework, in our case, the OECD guidelines. Table 4.2 is the complement of Table 4.1 presented in section 4.1; the intrinsic privacy characteristics are separated from those provided by the privacy model for a better understanding.

	OECD Directives							
	Collection Limitation	Data Quality	Purpose Specification	Use Limitation	Security Safeguards	Openness	Individual Participation	Accountability
FIA intrinsic privacy mechanisms								
Channel Security					■			
Message Security					■			
Pseudonymous				■				
Anonymous				■				
Usage Directives			■			■		
Interaction Services				■				
Privacy characteristics added by the model								
PI Management		▨					▨	▨
Privacy requirements			▨					▨
PI and policy access	▨	▨				▨	▨	▨
Policy management			▨			▨		▨
Policy enforcement				▨	▨			▨
Privacy preferences				▨				▨
Auditing module								▨

Table 4.2 Privacy management model and its relation with privacy principles

Table 4.2 shows the functionalities of the different modules from the privacy model proposed and how they could complement the fulfillment of the privacy principles of the OECD.

The intrinsic privacy mechanisms of the FIA are mainly focused to guarantee privacy during the attribute interchange between the *Attribute Provider* and the *Attribute Requester*, whereas the proposed privacy model enhances privacy by empowering users and entities to specify and enforce policies. The model emphasizes the logging of most events in order to facilitate the auditing process.

The collection of personal information is controlled and limited by users through the interface that allows them to introduce their personal data. Privacy specifications from organizations and external legal frameworks at the AP's side, allow users to know how their personal information is going to be treated by the AP, fulfilling in this way the privacy principles of "Purpose Specification" and "Openness". Policy access control provides the protection of personal information by enforcing the policies and disclosing the personal information only to the entities and for the purposes previously established, therefore it constitutes one of the most basic security safeguards. Access control fulfills directly the "Use Limitation" and "Security Safeguards" privacy principles. Finally, the log functionality proposed within the model guarantees the "Accountability" principle by allowing users and privacy officers to know how the personal information is handled during its collection, processing, storage and transmission.

The implementation of all the functionalities of the privacy model enhances substantially the level of privacy of the Federated Identity Architecture during the process of attribute sharing.

4.4 Conclusions

The privacy model proposed targets to enhance the level of privacy of the FIA system by supporting technologically a regulatory framework. The model allows protecting personal information with privacy policies which are built with external and internal regulations, as well as privacy preferences of the user. The policies are not only defined, they are enforced for each attribute access, whenever the personal information is collected, consulted or modified.

The model presents functional strengths that enhance to some extent the privacy within the FIA; however, it may add some challenges in terms of performance and scalability. The detected strengths and challenges are described in the following sections.

Among the detected strengths, we can mention:

- The proposed architecture is a 3-layer model with well defined components' functionalities; therefore, the architecture of each component is implementation independent.
- The model may work with any privacy regulatory framework, either, local, national or international.
- It is based on standardized web services and data structures. This guarantees the interoperability of solutions built with heterogeneous technology.
- The functionality of the components can be centralized or distributed within the CoT, providing flexibility at the development and implementation stages.
- The model is centered on user´s privacy rights and regulatory compliance.

Among the detected challenges, we can mention:

- The model logs all the activities, but it does not define any precise auditing processing.
- It is necessary to identify and prevent the possible security vulnerabilities introduced by the new functionalities of the model (if any).
- The model might have performance problems due to the great volume of transactions for a complex and demanding architecture.
- It could present scalability issues when the model is extended to support inter CoTs interactions.

4.5 Conclusions (en français)

Le modèle de respect de la vie privée proposé, vise à mieux respecter la vie privée dans une architecture d´identité fédérée en soutenant technologiquement un cadre légal. Le modèle permet la protection des informations personnelles avec les politiques en matière de respect de la vie privée qui sont établies avec des règlements externes et internes, aussi bien que des préférences de respect de la vie privée de l'utilisateur. Les politiques sont non seulement définies, elles sont imposées pour chaque accès d'attribut, toutes les fois que les informations personnelles sont rassemblées, consultées ou modifiées. Le modèle présente les avantages de fonctionnalité qui augmentent dans une certaine mesure le respect de la vie privée de la FIA ; cependant, il ajoute également quelques défis en termes d'exécution et efficacité. Ces défis représentent une amélioration continue du modèle. Quelques avantages et défis sont présentés ci-dessous.

Parmi les avantages détectés, nous pouvons citer:

- L'architecture proposée est un modèle en 3 couches avec les fonctionnalités des composants bien définies; par conséquent, l'architecture de chaque composant est indépendante de sa fonction.
- Le modèle fonctionne avec n'importe quel cadre de normalisation associé au respect de la vie privée, qu'il soit local au pays, national ou international.
- Il se base sur la normalisation du « Web Services » et des structures de données. Ceci garantit l'interopérabilité des solutions fondées sur des technologies hétérogènes.
- La fonctionnalité des composants peut être centralisée ou distribuée dans le cercle de confiance, ce qui fournit une grande flexibilité de développement et de mise en œuvre.
- Le modèle est centré sur le respect de la vie privée de l'utilisateur en conformité avec la normalisation.

Parmi les défis détectés, nous pouvons citer:

- Le modèle enregistre toutes les activités, mais il ne définit aucun processus d'audit.
- Il est nécessaire d'identifier et d'empêcher toute nouvelle vulnérabilité de sécurité qui pourrait être introduite par les nouvelles fonctionnalités du modèle.
- Le modèle peut connaître des problèmes d'exécution dus au grand volume de transactions d'une architecture complexe.
- Il peut présenter des problèmes de passage à l'échelle, en particulier si le modèle est étendu aux interactions entre cercles de confiance.

Chapter Five

Case scenario: Privacy model for Mexican e-Government services

Federated identity architecture has become a suitable solution for identity management within collaborative environments where entities must deploy services in a fast, secure and efficient way. The flexible grouping of entities in circles of trust facilitates the implementation of identity services such as single sign on and personal information sharing. However, this could represent a privacy risk if the inherent privacy mechanisms of FIA are not reinforced by additional functionalities in fulfillment with a privacy legal framework.

This chapter presents a case scenario for the deployment of e-Government services within the context of the mexican project known as e-Mexico National System. The integration of a federated identity architecture with the privacy model presented in chapter four is proposed as a solution for enhancing the management of citizens' digital identities as well as the improvement of privacy requirements compliance.

The first part of the chapter describes the benefits of electronic services to citizens and government entities in terms of cost reduction, service quality and fast deployment; the e-Mexico project is presented as an outstanding effort to bring information, knowledge and services in four important areas: education, health, commerce and government. The current handling of citizens' personal information by the federal public administration is described regarding administrative, technical and legal mechanisms for personal data protection. Next, a federated identity architecture is proposed for building a circle of trust composed of government entities so they can implement single sign on and attribute sharing functionalities. Finally, a step by step process is developed in order to implement the privacy model within the project of e-government services deployment that fulfills the mexican regulatory framework regarding privacy.

5.1 e-Government context

The growing demand of new government services needs a secure and fast sharing of sensitive information among government, organizations, business entities and citizens. Governments are often the source of personal information that relates to citizen´s identity such as birth certificates, driver´s license, tax records, marriage and death certificates among others (Candia, 2004).

As governments seek to extend their relationships with citizens and organizations, they are challenged to grant access to services and applications to the right people at the right time without compromising privacy, security or scalability. Nowadays, governments need to have a solution to meet the following requirements:

- To simplify citizen´s access to services.
- To reduce the cost and complexity of managing identities.
- To enable the dynamic creation of trusted relationships between different parties of the federal public administration.
- To share personal data preserving autonomy, privacy and ensure data security.

Governments around the world are promoting e-government initiatives such as the eEurope 2005 Action Plan and the e-Mexico National System project designed to foster the development of new and better services giving the citizens the opportunity to participate in the global information society.

A key element of the e-government strategy is the management of the citizen´s identity. The federated identity architecture represents a suitable solution as it allows government authorities to act as identity providers for citizens by establishing circles of trust and offering a complete range of personalized applications across different government organizations and domains.

The benefits of implementing a federated identity architecture within an e-government infrastructure can be described as follows:

a) Alliance improvement.- Enhances collaboration among government entities guaranteeing interoperability and preserving their autonomy in terms of personal information control.
b) Fast response time.- Creates a standard interface for identity services, making it easier to add and remove parties for critical communication services.
c) Cost reduction of services.- Increases individual and national productivity by granting citizens faster and easier access to applications and information throughout all agencies of the federal public administration.
d) Personal information security.- Provides gradient levels of authentication and risk management supporting different levels of sensitive services.
e) Easy integration.- Enables integration of identity management legacy systems without reengineering their authentication and authorization modules, because the federated identity architecture is built on standards.

There are a significant number of federated identity implementations around the world within the government sector. Such implementations take advantage from the most basic to the most complex features of the federated identity architecture depending on the e-government services maturity. Table 5.1 shows some examples of federated identity implementations within the government sector (Liberty, 2009).

Country	Federated Identity implementation	Population impact
Belgium	e-services for citizens and employees.	10 million
Denmark	SSO and federated services in the public sector.	3 million
France	Portal “Mon Service Public” which allows citizens to access a set of services.	60 million
New Zealand	Federated authentication system to support citizen and government interaction services.	4 million

Table 5.1 Examples of Federated Identity projects within the government sector

Next section describes the e-Mexico project started by mexican government as the main initiative to deploy electronic services to all mexican population.

5.2 e-Mexico National System project

On December the 1st, 2000, Mexico's President announced the beginning of the e-Mexico project, which has as commitments to bring more mexicans to information and knowledge through information and communication technologies, thus democratizing the access to information and services.

The e-Mexico National System is an integrating project which brings together the interests of several levels of government, of various public entities and divisions, of the telecommunication network operators, of the chambers and associations linked to Information and Communications Technology (ICT), as well as some other organizations, with the purpose of expanding the coverage of basic services in education, health, economy, government, science, technology and industry, as well as other services for the community (eMexico, 2009).

The strategy designed to implement the e-Mexico National System as a national program to reach the information and knowledge society, was divided into three main action lines:

Connectivity.- Its is focused on the investments being made by the operators of the telecommunication networks to increase the infrastructure and coverage of Internet access in mexican homes and the creation of a network of Digital Community Centers (DCC). This action connects residences and families that, because of economic and geographic limitations do not have the telecommunication infrastructure necessary to allow dedicated connectivity within their residence.

Systems.- It concerns with the creation of synergies to develop services platforms, integrating efforts and facilities of different e-Mexico participants. In this way, several services platforms are developed such as call centers, electronic signature, portals development platforms among others.

Content.- The Mexican Federal Public Administration (MFPA) which is organized into Ministries, Institutes, Councils and Agencies assumes as one of its most important commitments to facilitate to the mexican people the access to information and knowledge using the information and communication technologies. The content of the system is divided into the following four fields:

1) ***e-Learning***.- It provides new options for Mexicans to access to knowledge, education and training. It helps promoting the education and culture that should be accessible for all the citizens, having in mind the respect for their identity and cultural environment.
2) ***e-Health***.- It makes accessible to the entire Mexican population general medical and social security information and services. This helps promoting the human development and the improvement of the health institutions. This eliminates the barriers to the access of the information and the health and social security services.
3) ***e-Commerce***.- The system accelerates the development of the digital economy within businesses, specially micro, small and medium size businesses. It increases the competitive position of the Mexican economy, as well as to contribute developing a digital culture within the society, particularly among the consumers.
4) ***e-Government***.- The system provides the means so that all mexicans at the federal, regional, and municipal level can exercise their right to be fully informed and to have access to government services regarding citizens´ identification information such as civil registration, nationality, immigration, voting rights, certificates, among others.

The content above mentioned can be accessed independently through different governmental portals. However, the e-Mexico system proposes a centralized portal that integrates, by means of a single window, and in a harmonious manner, the four initial sheds of the e-Mexico national system, also it allows access to the services provided by the public federal, state and municipal administrations. The portal is becoming a means that fosters the citizen participation and improve the state-to-society, society-to-society, or society-to-state relationships, in an efficient, transparent and secure manner 24 hours a day, 365 days a year.

The following section explains how personal information of mexican citizens is handled by the MFPA.

5.3 Personal information handling by the Mexican Federal Public Administration

In Mexico, the IFAI (Instituto Federal de Acceso a la Información) is an independent organism within the Federal Public Administration. IFAI is responsible for ensuring that sensitive information, such as personal data in the custody of the federal government, is protected. In this role, the Institute issues guidelines for protecting personal data and guidelines for releasing it to its rightful owner when requested (Sobel, 2006).

Table 5.2 shows how IFAI has classified personal information into twelve categories and a set of citizen´s attributes for each category (Persona, 2006). Additionally, IFAI has emitted a set of security measures (Basic, Medium and High) associated to personal information categories that must be applied to the technological elements that collect, process or transmit the corresponding attributes (IFAI2, 2003).

Personal Information Category (PIC)	Security level	Attributes	
1.-Identification data	Basic	Full name Date of birth Place of birth Nationality Age Marital status Maternal language Habits Photo Signature Family dependants´ name Gender	Address Digital signature Private phone number Private cellular phone number e-mail CURP (Clave Unica de Registro de Población) RFC (Registro Federal de Contribuyentes) Military ID number Other
2.- Labor data	Basic	Title job Job position Job address Enterprise e-mail Enterprise phone number Past title jobs Job recommendation letters	Personal recommendation letters Extracurricular activities Training certificates Recruitment documents Appointment letters Labor incidence documents Other
3.-Patrimonial information	Medium	Income and debits Bank accounts Insurances Credit history Fiscal information Credit references	Guarantees Real estate and personal properties Contracted services Individual retirement account Other
4.-Academic data	Medium	Academic trajectory	Professional distinctions

		Professional titles Professional identity number	Certificates Other
5.-Judicial procedures	Medium	Judicial history resolutions Other	
6.-Migratory transit	Medium	Migratory status International transit Other	
7.-Physical characteristics	High	Complexion Physical constitution Weight Height	Identification marks Iris color Hair color Other
8.-Personal characteristics	High	Blood group Fingerprint	DNA Other
9.-Origin information	High	Ethnic group Racial origin Other	
10.-Ideological information	High	Religion Membership to religious organization Political affiliation	Ideology Membership to civil organizations Syndical affiliation other
11.-Health information	High	Disabilities Diseases Medical treatments Auxiliary medical devices Allergies Health status	Clinical file Consumption of toxic substances Psychological information Surgical operations Medical insufficiencies Other
12.-Sexual life	High	Sexual preferences Sexual habits Other	

Table 5.2 Personal information categories as defined by IFAI

All information systems that handle personal information of citizens within organizations of the MFPA are named Personal Data Systems (PDS). IFAI has deployed an information system known as *Persona* that allows government entities to classify and to register their PDS. The *Persona* system also allows citizens to verify the information systems managed by the government entities and the personal data handled by such systems. The *Persona* system represents an outstanding effort from the mexican government to provide transparency to mexican citizens regarding their personal data handling.

Despite the activities carried out by IFAI and government entities concerning personal information handling and transparency, there are some missing issues that must be considered:

a) Each government entity has its own citizen´s identity information which is not related or linked. Personal information of the same citizen is probably duplicated and there is no definition of which entity is authoritative (official source) for such personal data.
b) There is no standard specification of the syntax and semantics of personal data. The government entities exchanging personal information must build their own interfaces to communicate their PDS.
c) Most of the PDS do not allow citizens to access their personal information in order to verify or update it.
d) The citizens are not able to express their privacy preferences; therefore, they do not know the purpose of the collection, process and transmission of their personal data (see chapter four).

e) The privacy policies (if they exist) that some government entities present to citizens, generally are sentences that explain the intended use of personal information. However, there are no technical mechanisms to enforce the privacy policies.
f) Government entities must report to IFAI any transfer of personal data between PDS; however, there is no automatic mechanism to audit the compliance with privacy regulatory frameworks.

The federated identity management represents a suitable solution for issue (a) where the government entities can build circles of trust in order to federate the identities and exchange personal data of citizens in a secure way. The privacy model described in chapter four provides the necessary elements to fulfill the requirements exposed in issues from (b) to (f).

5.3.1 Mexican e-Government legal framework

The federal public administration must comply with the Federal Law of Transparency and Access to Government Public Information enacted in 2002 by IFAI (IFAI3, 2002). This law is founded in article 6 of the Mexican Political Constitution which establishes that all mexican citizens have the right to access the public information handled by the entities of the federal public administration. The law of transparency represents the first legal effort to allow the citizens know what personal information is handled by a particular government entity, and it also represents the right to access their own personal data. Although the law is focused on transparency, it guarantees in some extent the privacy of personal data because the government entities are enforced to guarantee its integrity, confidentiality and availability.

IFAI enacted in September 2005 the Guidelines for Personal Data Protection (IFAI, 2005). This guideline provides the government entities a set of privacy principles for protecting and guaranteeing privacy of personal data handled by the federal public administration. The guideline is compatible in some extent with the privacy principles proposed by the OECD, except for the last principle of each guideline (Transmission and Accountability respectively).

Following are the seven privacy principles specified by IFAI:

1. **Lawfulness**.- Personal data collected by government entities must be obtained only through legal means, and it must be used only for the specified purpose.
2. **Data quality**.- Personal data treatment must be accurate, adequate and not excessive in relation to the legal attributions of government entities.
3. **Access**.- Personal data systems must allow citizens to access their personal information.
4. **Notification**.- The collection and use purpose of personal data must be notified by government entities to citizens.
5. **Security**.- Government entities must implement the security mechanisms to guarantee the integrity, confidentiality and availability of personal data.
6. **Custody**.- Government entities must guarantee a secure treatment of personal data under their custody.
7. **Transmission**.- The transmission of personal data must have the consent of its owner.

Currently, two proposals for the constitutional amendment have been approved (Senate, 2008). The first, presented before the Senate, adds several paragraphs to article 16 of the federal constitution, explicitly acknowledging the right to personal data protection as a fundamental right. The proposal passed the Senate during the last legislative session, and was introduced before the House of Representatives. The amendment of article 16 establishes:

"Article 16.- All people has the right to the protection of his personal data, to the access, rectification and cancelation of it, as well as to declare his opposition, in the terms that the law determines, which will establish the exception assumptions by reasons of national security, public order disposition, public health and security or to protect the rights of third parties."

The second proposal adds the fraction XXIX-O to article 73 of the federal constitution which establishes:

"Article 73.- The Congress has the faculty:
...
XXIX-O. To legislate on personal data protection held by private entities.
..."

Additionally to the legal framework development, mexican government has started the deployment of a single identity identifier as a key element to simplify the assignment and management of digital identities of citizens. The following section describes the characteristics of such identifier.

5.3.2 Citizen single ID identifier

Historically, the government entities have assigned a different identification number to citizens for each specific sector, for example: social security number (SSN), passport ID, tax number, electoral ID, among other. However, actually many governments have started initiatives to develop identity management systems that include the implementation of a Single Identification Number (SIN) as exemplified in the report of Interoperability of eGovernment systems of the European Union Council (Lippmann, 2005).

The report presents the most important aspects to consider when implementing a national SIN: responsible organizations of assignment and monitoring, legislative framework, SIN allocation, personal information linked to the SIN, database sharing and technical construction of the identifier.

The MFPA started in the middle of the 90´s a project for the deployment of a single identification number known as CURP (Clave Unica de Registro de Población). The entity responsible for CURP assignment is RENAPO (Registro Nacional de Población) an agency of the Ministry of Interior; on the other hand, IFAI is the organization responsible for monitoring the legal framework compliance that was described in section 5.3.1.

The CURP is assigned to all Mexican citizens living in national territory or abroad and it has associated the necessary personal information to give citizens a legal identity (full name, gender, place of birth, date of birth and nationality). The CURP is stored in a central database managed by RENAPO which can be accessed by any entity government that complies with technical requirements and legal agreements. The following section describes the main characteristics and the structure of CURP.

Technical construction of CURP

Basically, there are two different approaches to construct the SIN. The first and bigger group of countries uses a semantic approach in terms of coding personal information within the identifier. The second group uses random numbers, where data protection is an important concern.

In the mexican case, the CURP is an alphabetic code (semantic approach) with the following characteristics: it is 18 alphanumeric characters long; it assures a biunivocal correspondence between the code and people; it is auto-generated from the personal data of the citizen such as name, gender, date and place of birth.

The structure of the code is as follows (Segob, 2006):

Characters 1 to 4: Initial letter and first internal vowel of the surname, the initial letter of the mother´s maiden name and the first letter of the first name.
Characters 5 to 10: Date of birth with yy/mm/dd format.
Character 11: Gender; M for female and H for male.
Characters 12 to 13: Initial letter and last consonant of the birthplace State.
Characters 14 to 16: First internal consonants of surname, mother´s maiden name and first name.
Character 17: Sequential character to differentiate possible homonymies.
Character 18: Verifier digit.

Figure 5.1 illustrates the CURP structure for a man with name Roberto Morales Ramos, who was born on the 20th of June 1960 in Mexico, DF.

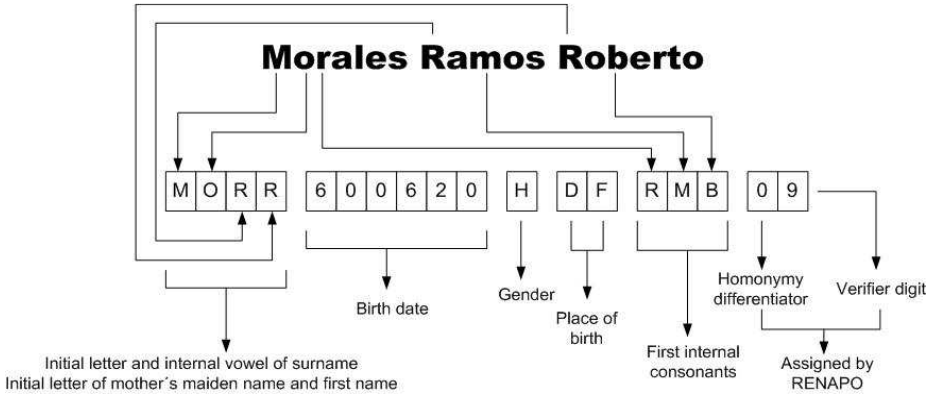


Figure 5.1 Structure of the CURP

Now, the CURP is widely used in the identification documents within the federal public administration, and it has become a key element in the identity projects such as the Unique Identity Card (UIC) and citizen identity management systems.

Once the current handling of personal information by the federal public administration has been presented, the following sections explain how the federated identity architecture and the privacy model can be integrated for deploying e-government services guaranteeing a high level of privacy.

5.4 e-Government and Federated Identity Architecture

This section proposes how the entities from the federal public administration can be organized into CoTs by sectors as classified in the e-Mexico national system project Figure 5.2 shows a possible federated identity architecture composed of four basic CoTs (e-Learning CoT, e-Health CoT, e-Commerce CoT and e-Government CoT) interacting among them and with other CoTs of non-governmental sectors. All the CoTs need establishing trust relationships (T) among them, allowing the inter-CoT communication for SSO and attributes exchange functionalities.

A detailed structure for the e-Government CoT is presented in Figure 5.2. Each entity within the circle of trust manages the identity information of its users (ID1, ID2 and ID3) and provides a certain type of services and information, that is, they can play the role of IdP or SP. The identities are federated through pseudonyms (F1, F2 and F3), so that a particular entity does not know the identity information managed by other entities. Additionally, a trust relationship exists among the entities composing the CoT allowing them to exchange personal information and implement Single Sign On and Single Log Out functionalities. The trust relationship consists in a common encryption infrastructure and business agreement.

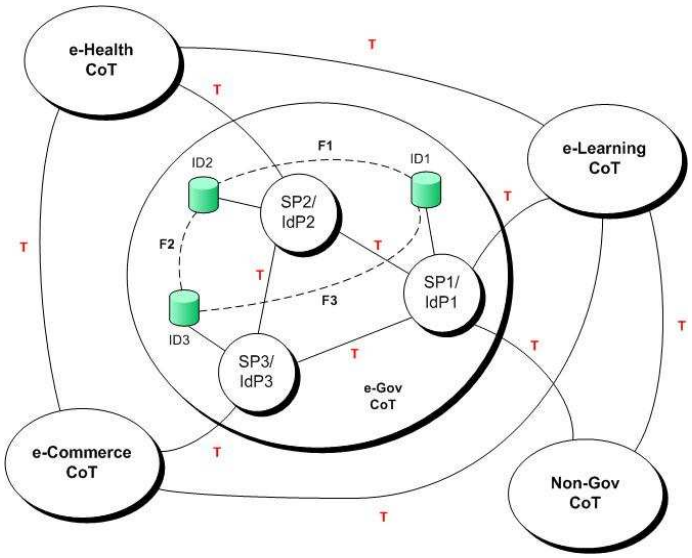


Figure 5.2 Possible architecture of CoTs for the MFPA

In a real environment, the entities composing a CoT require interacting with entities from other CoTs, thus there must be a trust relationship with one or more entities representing each CoT.

The architecture of the interconnected CoTs can be as complex as required by the services to be provided to the citizens. The facility to construct and interconnect the CoTs provides flexibility and scalability to the architecture.

The three entities of the e-Government CoT manage the citizens´ IDs shown in Figure 5.3. The first field within the digital identity represents the main identifier; it can be a username, a numeric ID or any attribute that identifies univocally to the citizen. Additionally, there are some personal attributes associated to the citizen, where the last one is assigned by the entity as a credential to carry out the authentication process. The credential can be a password, a Personal Identifier Number (PIN) or even a digital certificate.

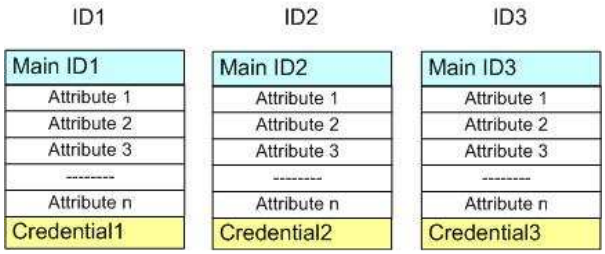


Figure 5.3 Digital identities managed by the IdPs of the e-Government CoT

The previous digital identities are federated through a pseudonym as shown in Table 5.3, it represents a mapping between the local identifier and the corresponding pseudonym. The federation is carried out by government organizations to implement the SSO and simplify the attribute exchange among the entities composing the CoT.

Federation between	Identifier 1	Identifier 2	Federated pseudonym
SP1/IdP1 – SP2/IdP2	Main ID1	Main ID2	F1
SP2/IdP2 – SP3/IdP3	Main ID2	Main ID3	F2
SP3/IdP3 – SP1/IdP1	Main ID3	Main ID1	F3

Table 5.3 Digital identities federation

With the federation of identities, functionalities such as single sign on can be implemented as shown in the following case scenario:

A mexican citizen authenticates at the e-Mexico portal (central portal hosted at SP1); however, the service required is provided by SP2. Figure 5.4 shows the process of SSO and service access within the e-government CoT.

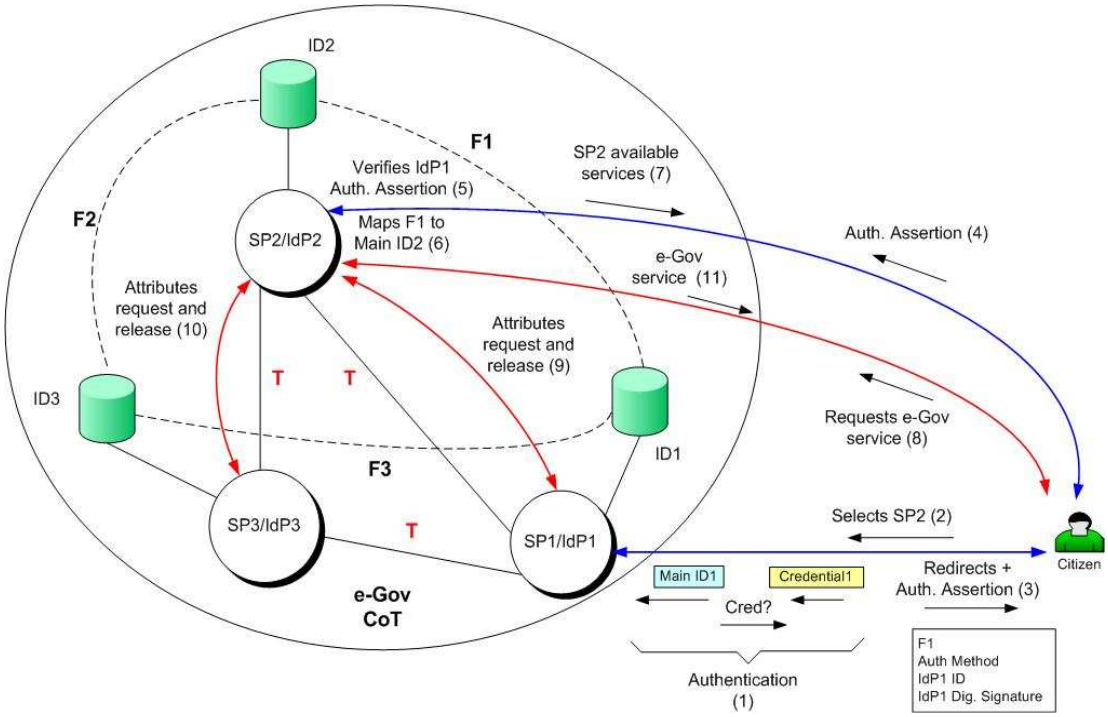


Figure 5.4 SSO and attributes exchange within e-Gov CoT

The citizen authenticates with the SP1/IdP1 portal which plays the role of IdP, giving his identifier Main ID1 and his Credential1 (1). The portal validates the identity allowing the citizen to access the services including those from other SPs; the citizen requests services provided by SP2 organization (2) and the IdP1 builds an Authentication Assertion (3) containing the following information: the pseudonym (F1), the authentication method used (Auth Method), the IdP1 identification, and its digital signature over the emitted assertion.

The citizen is redirected to the SP2 portal presenting the authentication assertion (4). The SP2 verifies the authenticity of the assertion´s digital signature (5) using a symmetric key or a PKI infrastructure depending on how the trust relationship was established. If the assertion is authentic, then the SP2 maps the F1 pseudonym to the local user with the corresponding Main ID2 (6) and then, the available services are presented to the citizen (7).

The citizen requests the corresponding e-Gov service (8) which may require additional personal information retrieved from IdP1 (9) or IdP3 (10). The attributes are requested using the pseudonyms F1 or F2 respectively. After the attributes are released to SP2, the information is processed in order to complete the e-Gov service (11).

As shown in the previous case scenario, the federated identity architecture simplifies the management of citizen´s digital identity and assures the exchange of attributes by means of messages encrypted and digitally signed. However, in order to improve most of the privacy issues exposed in section 5.3, the privacy model is integrated for the deployment of e-government services as presented in the following section.

5.5 Case scenario for e-Government services

5.5.1 E-Government service context

This case scenario presents the process for implementing the privacy model described in chapter four into the e-government service deployment.

A simplified scenario is presented where entities government EgovA and EgovB are members of a CoT and the digital identities of the citizens are federated providing identity services such as SSO and attribute exchange. EgovA plays the role of IdP and EgovB plays the role of SP. EgovA contains the Attribute Provider (AP) component which releases citizen´s personal information to Attribute Requester (AR) component of government entities requesting personal data. Privacy policies are associated to PI and they are enforced at the IdP side each time attributes are requested. Figure 5.5 shows the interactions among the citizen and government entities A and B when the citizen authenticates with EgovA and requests a service to EgovB.

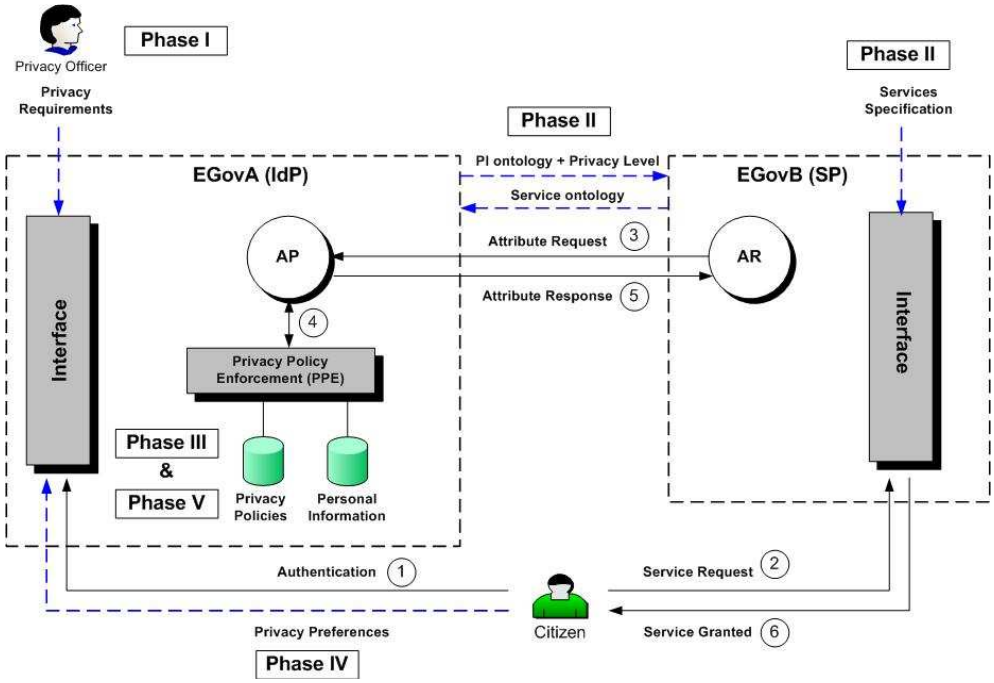


Figure 5.5 Case scenario of the privacy model for e-Government

The citizen authenticates with EgovA (1) and then requests services to EgovB (2); if the service to be provided needs additional personal data in order to be completed, it is requested (3) to the appropriate attribute provider (in this case is the same EgovA entity). The AP at EgovA side receives the request which contains: which attribute is needed, who is requesting the data, and the privacy parameters (purpose and retention time) according to the privacy level of the attribute. The AP analyzes the requesting parameters and then demands to the *Privacy Policy Enforcement (PPE)* module to evaluate the request (4). If the privacy policy is evaluated positively, the AP sends a response to the AR with the corresponding attributes; if the request did not comply with the privacy policy a denying response is sent (5). Finally, the service is granted to the citizen (6).

It is important to remark that for achieving all the above described functions, the privacy model must be deployed within the federated architecture. The methodology proposed for implementing the model is composed of five phases. Phase I includes the privacy requirements specification at the IdP side; phase II relates with the service definition at the SP side and the

ontology exchange between IdP and SP, this phase is important in order to match the services provided with the personal data required; phase III covers the privacy policy ontology generation; phase IV allows citizens to express their privacy preferences modifying the privacy levels for the particular attributes; phase V is the final step where privacy policies are built and ready to be enforced every time an attribute is requested. The following sections describe into detail each of the phases.

5.5.2 Privacy model implementation phases

The methodology is divided into the phases shown in Figure 5.6:

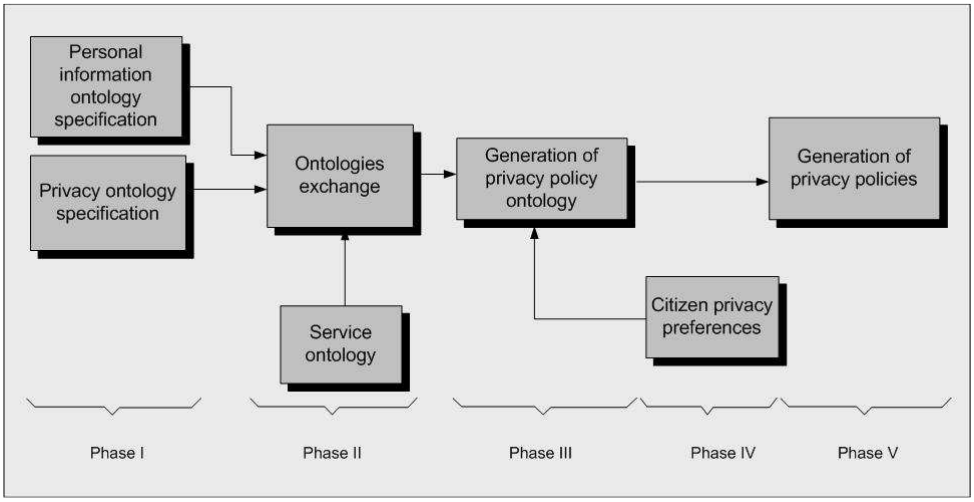


Figure 5.6 Privacy model implementation within the e-government services

Phase I: This phase is carried out by the government entities responsible for defining the citizens´ personal information ontology based on the PI classification made by IFAI as described in Table 5.2. Additionally, during this phase, experts (privacy officers) in the field of regulatory framework express the national privacy requirements in natural language. The privacy requirements specification is converted to formal expression through an ontology development methodology. In the case of Mexico, the regulatory framework is the Guidelines for Personal Data Protection (IFAI, 2005); additionally, I propose three privacy levels (Basic, Moderate and Strict) in accordance with the three security levels associated by IFAI to PI categories. The modules of the privacy model involved in this phase are: *Metadata Specification* and *PI Management* (see section 4.2.1 of chapter four).

Phase II.- Based on the personal information ontology specified in phase I, each government entity generates its own PI ontology subset only with the attributes that are handled by its personal data systems and for which this entity is authoritative (official provider). In this phase, the government entity sends its PI ontology (when playing the role of IdP) to entities acting as SPs who send in return the service ontology describing the services provided to citizens and the personal data required to complete them. The modules of the privacy model involved in this phase are: *Metadata Specification* (see section 4.2.1 of chapter four).

Phase III.- With the ontologies created and exchanged in phase I and phase II respectively, the IdP is able to construct the privacy policy ontology. This ontology combines the services that citizens are allowed to access with the personal information needed and the associated privacy levels. The module of the privacy model involved in this phase is: *Metadata Specification* (see section 4.2.1 of chapter four).

Phase IV.- The privacy policy ontology built in phase III represents the “default” privacy specification associated to each attribute and the corresponding services. At this point, citizens can express their privacy preferences, changing the privacy level for each attribute from low privacy level to high privacy level; that is, from basic to moderate or to strict level, or from

moderate to strict level, but not in opposite sense. The module of the privacy model involved in this phase is: *Personal Information and Policy Access* (see section 4.2.3 of chapter four).

Phase V.- After the citizen has expressed its privacy preferences; the privacy policy ontology is converted into privacy policy expressed in a standard language that allows the policy enforcement mechanism. The resulted privacy policy is stored and indexed with the corresponding attribute. The module of the privacy model involved in this phase is: *Privacy Policy Management* (see section 4.2.2 of chapter four).

The following sections detail each of the phases above described.

5.5.3 Phase I: Personal information abstraction and privacy requirements

5.5.3.1 Personal information ontology specification

The first step in the process is to generate the personal information ontology; Figure 5.7 shows the graphical representation of the complete ontology with categories and some attributes as defined by IFAI (Persona, 2006).

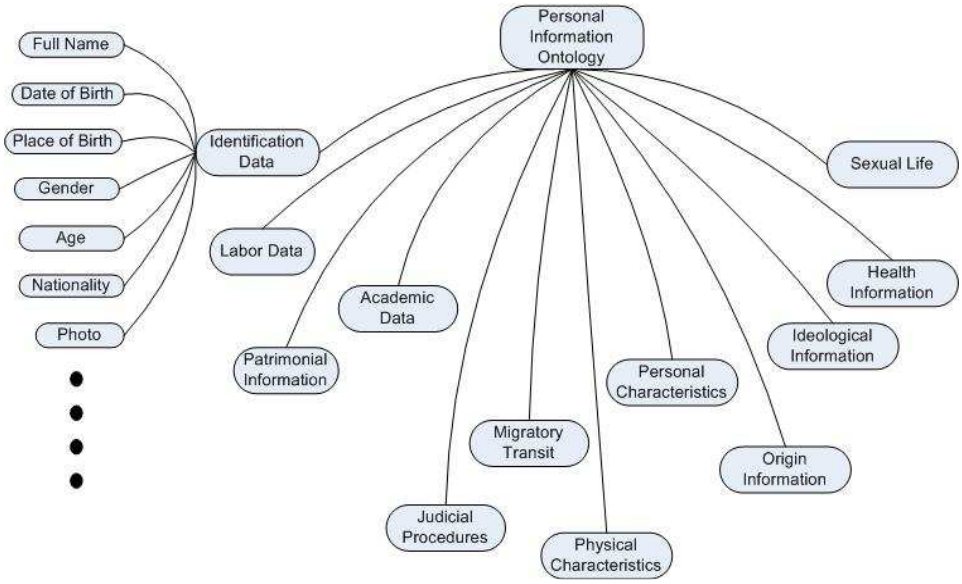


Figure 5.7 Graphical representation of PI ontology as defined by IFAI.

As explained is section 5.3, each attribute of personal data has associated a security level specifying the security measures that must be applied by all the technological elements that collect, process or transmit personal data. Similarly, a privacy level must be assigned to each attribute depending on the degree of privacy defined by the mexican government. For this case study, three privacy levels (Public, Moderate and Strict) are defined, but the number of privacy levels and the specification of each level depend on the privacy requirements for the particular government. Additionally, the attributes have a relationship with government entities indicating which entity is the authoritative provider for those attributes; the authoritativeness may be one, some, all or none. Figure 5.8 shows the relationship of personal information ontology with security and privacy specifications and government entities. The graphical is represented in ORM (Object-Role Modeling) which provides a way to express the meaning and relationship at conceptual level. The diagram is interpreted as: “PI has security level”, “PI has a privacy level”, and “Government entity is authoritative of some PI”.

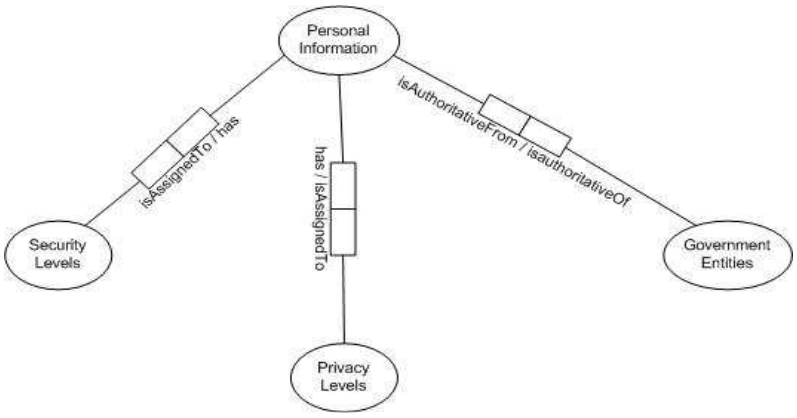


Figure 5.8 Security level, privacy level and authoritativeness associated with PI

For the construction of the ontology it is needed a tool that allows to specify the elements and relations of the ontology and that is able to generate the code in any ontology standard language. The software tool used to construct the ontologies is the Protégé editor which is an open-source platform developed by Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine (Protégé, 2009). Figure 5.9 shows the main screen of Protégé editor, where the personal information ontology is described as a tree of classes in Protégé terminology.

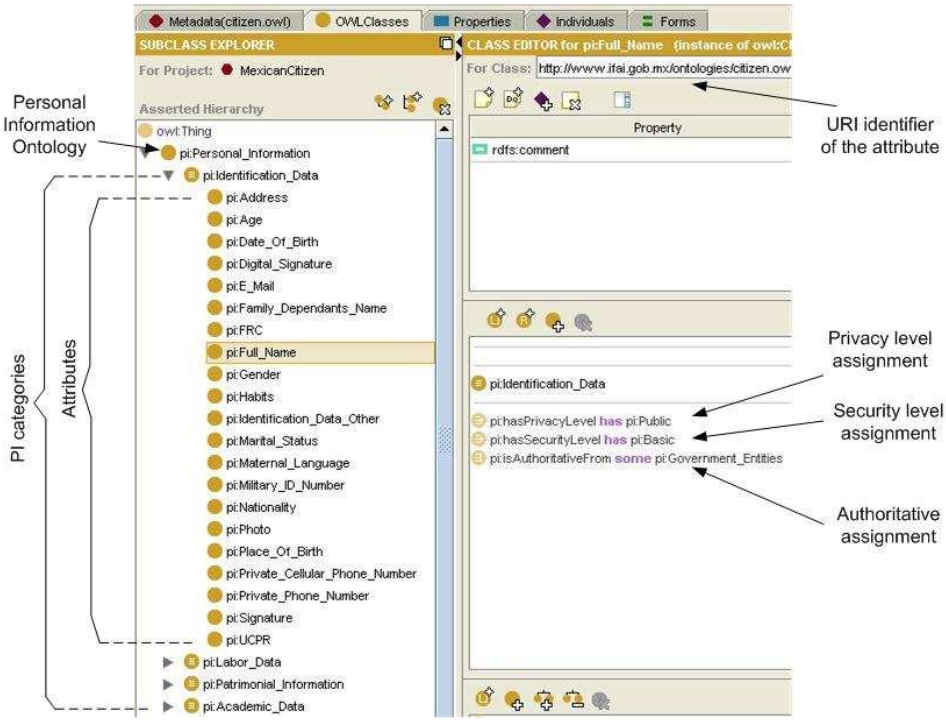


Figure 5.9 Personal information ontology specified in Protégé platform

The left side window shows the tree of objects (class in Protégé terminology) where **Personal_Information** object represents the PI ontology. The second level of the tree specifies the categories as defined by IFAI (ie. **Identificaton_Data**, **Labor_Data**, **Patrimonial_Information**, among others) and the third level specifies the individual attributes for each category of the attribute **Full_Name** is highlighted; therefore, the right part of the screen shows some of its characteristics. The item **hasPrivacyLevel** is a property that relates the attribute with **Public** privacy level; **hasSecurityLevel** property relates the attribute with the **Basic** security level value; meanwhile, the item **isAuthoritativeFrom** indicates that an element

of the class **Government_Entities** is the authoritative provider of such attribute. The URI identifier of the attribute guarantees its uniqueness.

Protégé ontologies can be exported into a variety of XML languages including RDF (Resource Description Framework) and OWL (Ontology Web Language); where RDF is a W3C standard which was designed to provide a common way to describe concepts so they can be read and understood by computer applications; meanwhile OWL is a family of knowledge representation languages developed by W3C for authoring ontologies. **¡Error! No se encuentra el origen de la referencia.** shows some fragments of the code generated for the attribute **Full_Name** of the PI ontology.

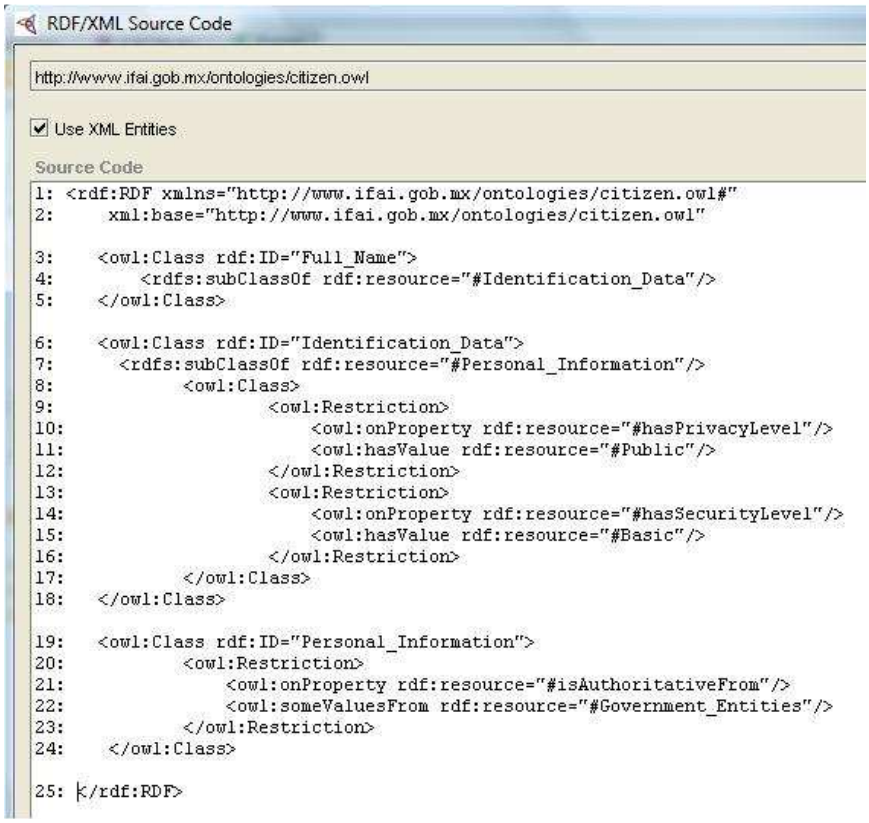


Figure 5.10 Fragment of the XML source code of personal information ontology

Lines 1 and 2 specify the name space for the ontology, in this case the URI corresponds to IFAI´s web site. Lines 3 to 5 define the attribute **Full_Name** as a subclass of the category **Identification_Data**. Lines 6 to 18 specify the category **Identification_Data** with restrictions **hasPrivacyLevel** and **hasSecurityLevel** with values **Public** and **Basic** respectively, such values are associated to categories and inherited by the attributes. This category is defined as a subclass of the ontology **Personal_Information** (line 7). Lines 19 to 24 define the ontology with the restriction **isAuthoritativeFrom** indicating that each attribute of the ontology has a government entity as authoritative provider.

5.5.3.2 Legal framework ontology and Privacy ontology specification

The process to translate the legal framework and privacy requirements expressed in natural language to formal expressions (ontologies) is carried out through a methodology developed by PRIME (Privacy and Identity Management for Europe). PRIME is a consortium of 20 member organizations from industry, academia, research centers and data protection established in 2004 focused on solutions for privacy-enhancing identity management that was partially funded by the European Commission under the 6th Framework Program (PRIME, 2009).

The methodology explained in the Ontology Development Process document (Tang, 2005) consists of three steps:

1. **Verbalize elementary facts**: from the source paragraph, the expert responsible of expressing the text can rephrase such text in order to segment it getting simple sentences in the form of *{subject-verb-object}*. This process is known as knowledge breakdown.
2. **Create lexons**: this activity uses the verbalize facts resulting from the previous step as input. The aim is to extract binary facts (or lexons) in the form of $\{concept_1, relation_{1\text{-}2}, relation_{2\text{-}1}, concept_2\}$, where *concept1* is the *subject*, *concept2* is the *object* and *relations* are the *verbs* that describe the relationship between them in both directions.
3. **Build ontology**: the lexons resulted from the previous step are used to build the ontology represented in ORM diagram and then converted into XML code.

The methodology is applied to build two ontologies: the Legal Framework Ontology (LFO) and the Privacy Ontology (PO). The LFO is used to express the privacy principles emitted by the IFAI; meanwhile, the PRO specifies the different privacy levels defined by the mexican government.

Legal framework ontology specification

Table 5.4 shows the verbalization and lexons creation steps applied to the privacy principles (see section 5.2.1) emitted by IFAI (IFAI, 2005). Each principle is segmented into several simplified sentences and each sentence is used to build the lexons.

1. Lawfulness	Personal data collected by government entities must be obtained only through legal means, and it must be used only for the specified purpose.	
Sentence 1.1	Government entities collect personal data	
	Lexon 1.1.1	{Government_Entities,collect,isCollectedBy,Personal_Data}
Sentence 1.2	Government entities must use legal means to collect personal data	
	Lexon 1.2.1	{Government_Entities,use,isUsedBy,Legal_Means}
Sentence 1.3	Personal data are used for specific purpose	
	Lexon 1.3.1	{Personal_Data,has,appliesTo,Purpose_Use}
2. Data quality	Personal data treatment must be accurate, adequate and not excessive in relation to the legal attributions of government entities.	
Sentence 2.1	Government entities must treat personal data accurately	
	Lexon 2.1.1	{Government_Entities,treat,isTreatmentOf,Accurately}
Sentence 2.2	Government entities must treat personal data adequately	
	Lexon 2.2.1	{Government_Entities,treat,isTreatmentOf,Adequately}
Sentence 2.3	Government entities must not treat personal data excessively	
	Lexon 2.3.1	{Government_Entities,treat,isTreatmentOf,No_Excessively}
3. Access	Personal data systems must allow citizens to access their personal information.	
Sentence 3.1	Personal data systems are used to access personal data	
	Lexon	{Personal_Data_Systems,areUsedToAccess,isAccessedThrough

	3.1.1	,Personal_Data}
Sentence 3.2	Citizens must be allowed to access their personal data	
	Lexon 3.2.1	{Citizens,access,isAccessedBy,Personal_Data}
4. Notification	The collection and use purpose of personal data must be notified by government entities to citizens.	
Sentence 4.1	Personal data have a purpose collection	
	Lexon 4.1.1	{Personal_Data,has,appliesTo,Purpose_Collection}
Sentence 4.2	Government entities must notify the purpose collection	
	Lexon 4.2.1	{Government_Entities,notify,isNotifiedBy,Purpose_Collection}
Sentence 4.3	Government entities must notify the purpose use	
	Lexon 4.3.1	{Government_Entities,notify,isNotifiedByPurpose_Use}
5. Security	Government entities must implement security mechanisms to guarantee integrity, confidentiality and availability of personal data.	
Sentence 5.1	Government entities must implement security mechanisms	
	Lexon 5.1.1	{Government_Entities,implement,areImplemented,Security_Mechanisms}
Sentence 5.2	Security mechanisms must be implemented to protect personal data	
	Lexon 5.2.1	{Security_Mechanisms,protect,isProtectedBy,Personal_Data}
Sentence 5.3	Security mechanisms must guarantee integrity of personal data	
	Lexon 5.3.1	{Security_Mechanisms,guarantee,isGuaranteedBy,Integrity}
Sentence 5.4	Security mechanisms must guarantee confidentiality of personal data	
	Lexon 5.4.1	{Security_Mechanisms,guarantee,isGuaranteedBy,Confidentiality}
Sentence 5.5	Security mechanisms must guarantee availability of personal data	
	Lexon 5.5.1	{Security_Mechanisms,guarantee,isGuaranteedBy,Availability}
6. Custody	Government entities must guarantee a secure treatment of personal data under their custody.	
Sentence 6.1	Government entities must treat personal data securely	
	Lexon 6.1.1	{Government_Entities,treat,isTreatmentOf,Securely}
Sentence 6.2	Government entities keep under their custody personal data	
	Lexon 6.2.1	{Government_Entity,handleCustodyOf,isCustodiedFrom,Personal_Data}
7. Transmission	The transmission of personal data must have the consent of its owner.	
Sentence 7.1	Government entities transmit personal data	
	Lexon 7.1.1	{Government_Entity,transmit,isTransmittedBy,Personal_Data}
Sentence 7.2	Citizen must consent transmission of his personal data	
	Lexon 7.2.1	{Citizen,consent,isConsentedBy,Transmission}

Table 5.4 Verbalizing and lexons creation from IFAI´s privacy principles

The elements (concepts) extracted from lexons composing the LFO are shown in Figure 5.11 “Privacy Principles” represent the seven principles specified by IFAI. “Subjects and Objects” relate the entities from MFPA, citizens, personal information and information systems that process personal data. “Purposes” define what the personal data are used and collected for. “Legal Collection Means” represent any official mean provided by government entities in order to collect personal data. “Security Mechanisms” made reference to security measures applied to all elements that handle personal data. “Personal Data Treatment” defines how personal data must be treated when it is handled by government entities. “Personal Data Properties” represent characteristics that must be guaranteed by government entities when processing personal data.

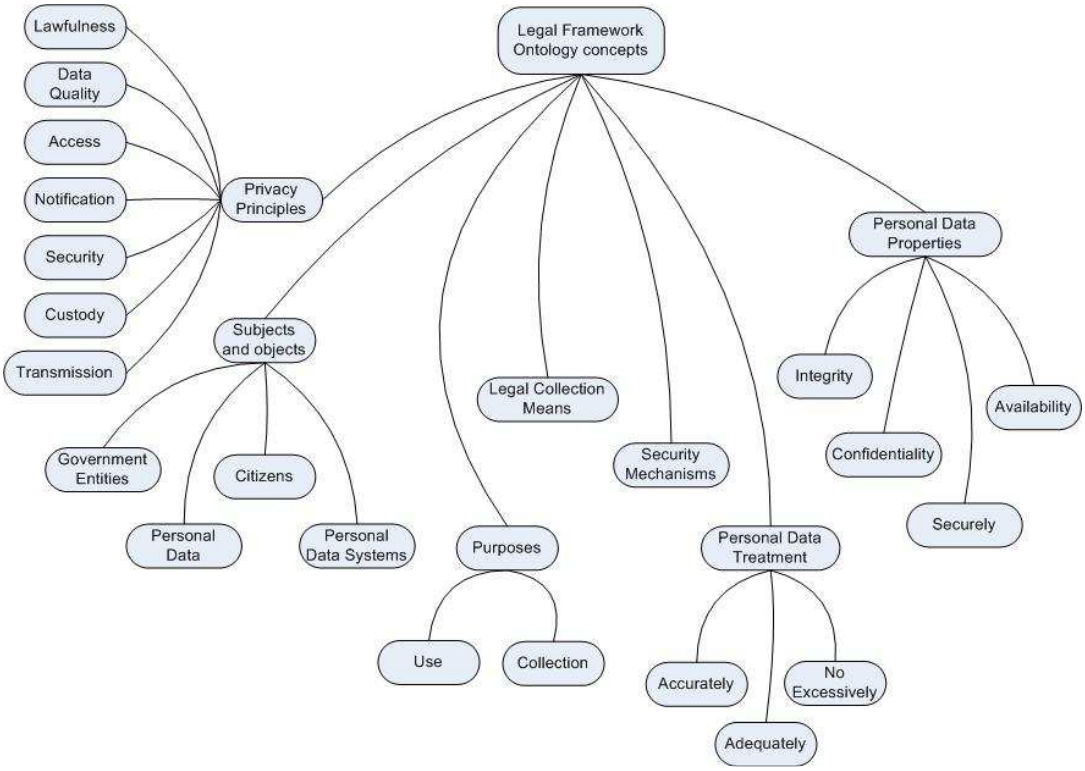


Figure 5.11 Legal framework ontology

The relations between concepts of the LFO are represented in the ORM diagram in Figure 5.12. The relation highlighted is interpreted in both senses as: “Government entities collect personal data” and “Personal data are collected by government entities”. This corresponds to lexon 1.1.1 of Table 5.4.

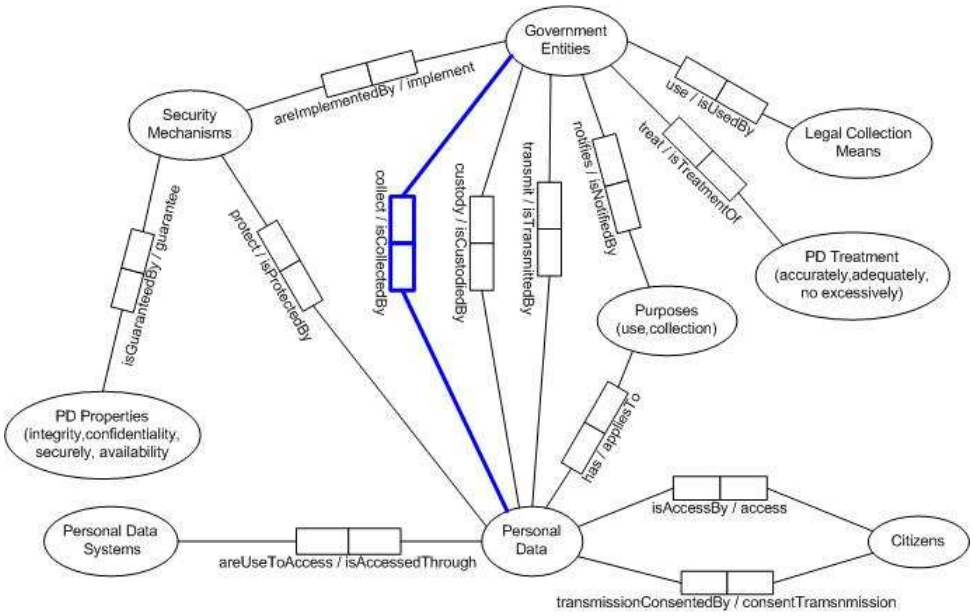


Figure 5.12 ORM diagram of the legal framework ontology

Finally, the legal framework ontology can be represented in Protégé platform. The left side window of Figure 5.13 shows the concepts specification; meanwhile the right central window describes the relations for the highlighted concepts (**Government_Entities**).

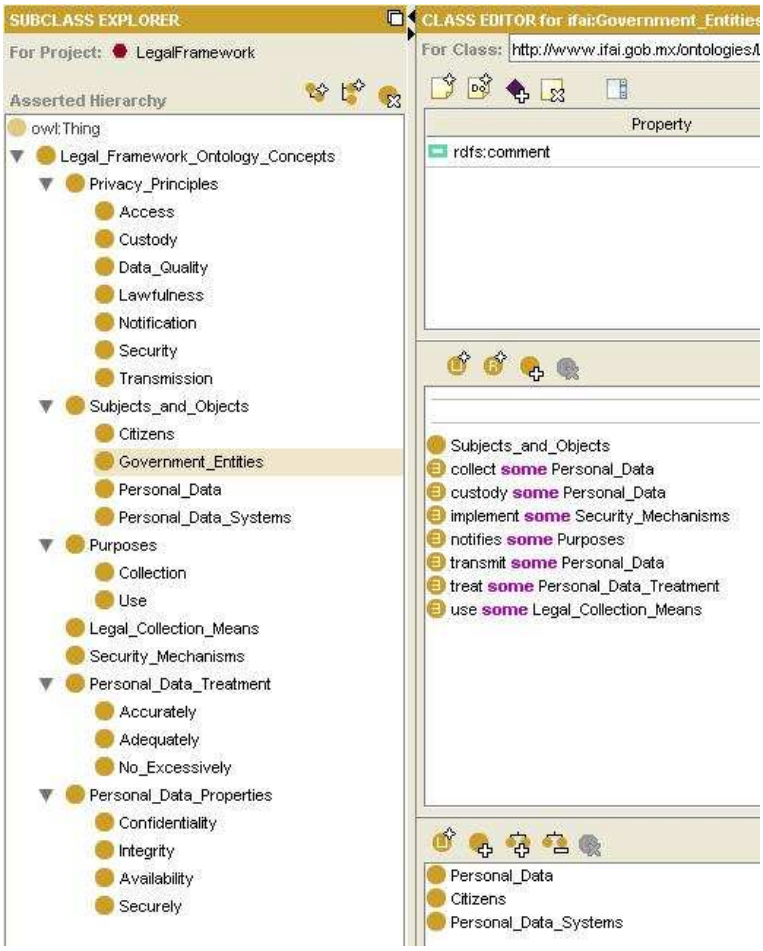
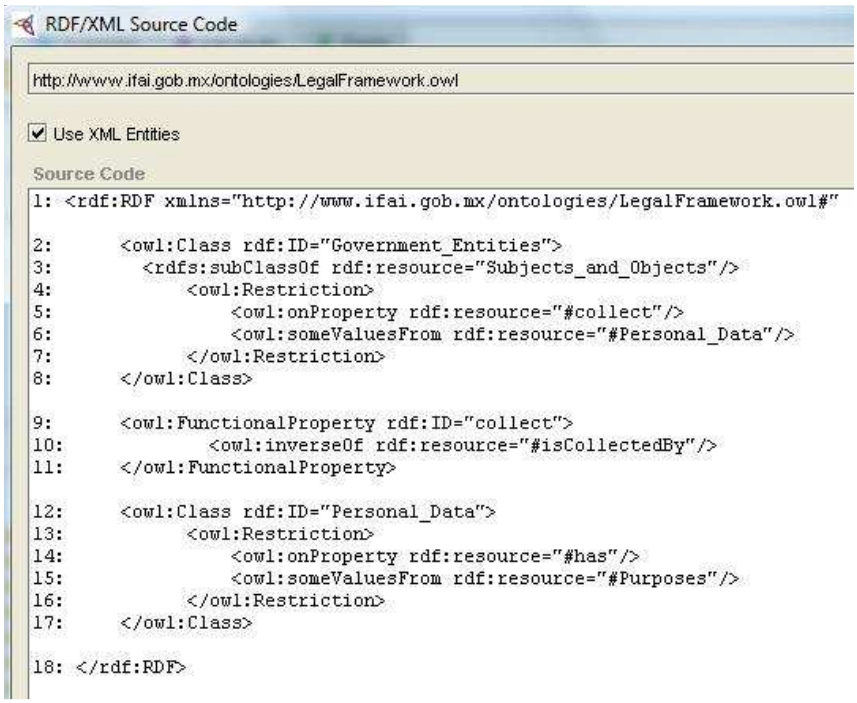


Figure 5.13 Legal framework ontology specified in Protégé platform

The XML code for the ontology can be generated as shown in Figure 5.14. The fragment code in OWL/RDF language describes the lexon 1.1.1 of Table 5.4. Line 1 specifies the URI value for the ontology. Lines 2 to 8 define **Government_Entities** as a subclass of **Subjects_and_Objects** with relation "collect personal data". Lines 9 to 11 define the relation **collect** as inverse of **isCollectedBy**. Finally, lines 12 to 17 specify that **Personal_Data** has a purpose value.

The LFO is not used to build the privacy policies; however, it defines in a formal language the legal requirements expressed in the privacy principles emitted by IFAI. This can be useful when the legal framework needs to be compared with others regulatory frameworks, mainly when the interaction is with other countries (functionality that is out of scope of the present work).


RDF/XML Source Code

http://www.ifai.gob.mx/ontologies/LegalFramework.owl

Use XML Entities

Source Code

```
1: <rdf:RDF xmlns="http://www.ifai.gob.mx/ontologies/LegalFramework.owl#"

2:      <owl:Class rdf:ID="Government_Entities">
3:        <rdfs:subClassOf rdf:resource="Subjects_and_Objects"/>
4:            <owl:Restriction>
5:                <owl:onProperty rdf:resource="#collect"/>
6:                <owl:someValuesFrom rdf:resource="#Personal_Data"/>
7:            </owl:Restriction>
8:      </owl:Class>

9:      <owl:FunctionalProperty rdf:ID="collect">
10:             <owl:inverseOf rdf:resource="#isCollectedBy"/>
11:     </owl:FunctionalProperty>

12:     <owl:Class rdf:ID="Personal_Data">
13:           <owl:Restriction>
14:               <owl:onProperty rdf:resource="#has"/>
15:               <owl:someValuesFrom rdf:resource="#Purposes"/>
16:           </owl:Restriction>
17:     </owl:Class>

18: </rdf:RDF>
```

Figure 5.14 Fragment of the XML source code of the legal framework ontology

Privacy ontology specification

The privacy ontology (PRO) is used to build the privacy policies along with the personal information ontology and services ontology. PRO specifies different levels of privacy that must be associated to each attribute of the citizen´s personal data. For this case study, I propose three privacy levels in order to be compatible with the security levels defined by IFAI; however, the number of privacy levels and their specifications can vary depending on the government requirements.

The proposed privacy level classification is: public, moderate and strict. Each level is expressed in natural language taking into account the privacy elements that answer *who is accessing the data (Recipient)?*, *what are the data used for (Purpose)?* and *how long the data will be kept (Retention)?*.

The privacy ontology is constructed using the PRIME methodology as described for the LFO; the natural language definition for each privacy level is:

Public.- Personal data are shared with any requesting entity whose usage privacy practices are not known and it may be used for any purpose. Personal data may be kept indefinitely or as long as permitted by the law.

Moderate.- Personal data are shared only with government entities with equivalent privacy practices and it may be used to personalize the services. Data may be kept for all the duration of the service provisioning or as permitted by the law.

Strict.- Personal data are not shared with other entities and they must be used only for the specified purpose. Data are kept only for all the duration of the service transaction or as required by the law.

Applying the verbalizing and lexon creation steps to the privacy levels specification, the result is shown in Table 5.5.

1. Public	Personal data are shared with any requesting entity whose usage privacy practices are not known and it may be used for any purpose. Personal data may be kept indefinitely or as long as permitted by the law.	
Sentence 1.1	Public personal data are shared with any entity.	
	Lexon 1.1.1	{Public_PI, areSharedWith,hasAccessTo,Any_Entity}
Sentence 1.2	Public personal data are used for any purpose.	
	Lexon 1.2.1	{Public_PI,havePurposeUse,isPurposeUseOf,Any_Purpose}
Sentence 1.3	Public personal data are kept indefinitely.	
	Lexon 1.3.1	{Public_PI,areRetained,isRetentionOf,indefinitely}
Sentence 1.4	Public personal data are kept as long as permitted by the law.	
	Lexon 1.4.1	{Public_PI,areRetained,isRetentionOf,Law_Permits}
2. Moderate	Personal data are shared only with government entities with equivalent privacy practices and it may be used to personalize the services. Data may be kept for all the duration of the service provisioning or as permitted by the law.	
Sentence 2.1	Moderate personal data are only shared with the government entities.	
	Lexon 2.1.1	{Moderate_PI,areSharedWith,hasAccessTo,Any_Government_Entity}
Sentence 2.2	Moderate personal data are used to personalize the services.	
	Lexon 2.2.1	{Moderate_PI,havePurposeUse,isPurposeUseOf,Personalization_Purpose}
Sentence 2.3	Moderate personal data are kept as long as the service is active.	
	Lexon 2.3.1	{Moderate_PI,areRetained,isRetentionOf,Service_Active}
Sentence 2.4	Moderate personal data are kept as long as permitted by the law.	
	Lexon 2.4.1	{Moderate_PI,areRetained,isRetentionOf,Law_Permits}
3. Strict	Personal data are not shared with other entities and they must be used only for the specified purpose. Data are kept only for all the duration of the service transaction or as required by the law.	
Sentence 3.1	Strict personal data are not shared with the third entities.	
	Lexon 3.1.1	{Strict_PI,areSharedWith,hasAccessTo,This_Government_Entity}
Sentence 3.2	Strict personal data are used for the specified purpose.	
	Lexon 3.2.1	{Strict_PI,havePurposeUse,isPurposeOf,Specified_Purpose}
Sentence 3.3	Strict personal data are kept for all the duration of the service transaction.	
	Lexon 3.3.1	{Strict_PI,areRetained,isRetentionOf,Service_Transaction}
Sentence 3.4	Strict personal data are kept as long as permitted by the law.	
	Lexon 3.4.1	{Strict_PI,areRetained,isRetentionOf,Law_Requires}

Table 5.5 Verbalizing and lexons creation from the privacy levels specification

The concepts composing the privacy ontology are shown in Figure 5.15. “Privacy Levels” define the three privacy levels values proposed to be assigned to each attribute. “Recipients” are the entities who have access to personal data; “This Government Entity” is the entity acting as an SP that requires the attribute (not shared with any other entity); “Any Government Entity” is any entity from the federal public administration that requires an attribute although it is not providing the service directly; “Any Entity” can be a non government entity. “Purposes” specify what the data are used for; “Specified Purpose” is when attribute is used only to complete the activity for which it was provided; “Personalization Purpose” is focused to tailor the service for each citizen; “Any Purpose” does not specify a particular use. “Retention” has different values: “Indefinitely” has no limit of time; “Service Active” indicates that attribute may be kept as long as the SP is providing the service; “Service Transaction” specifies that the attribute must be discarded after each transaction. “Law Requires” and “Law Permits” are periods of time defined by the national regulations.

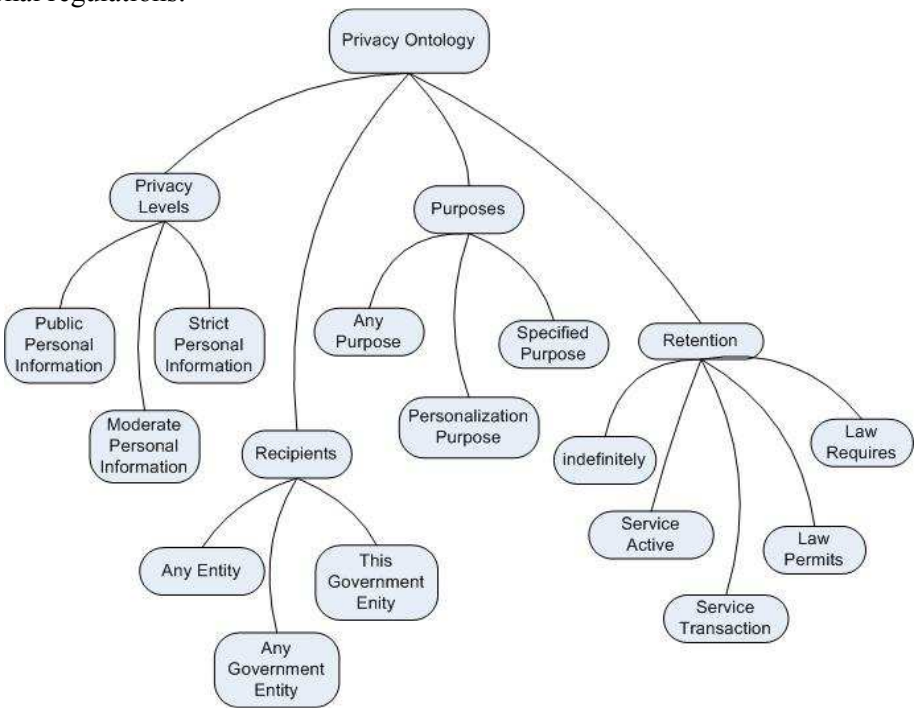


Figure 5.15 Concepts composing the privacy ontology

The relationship between the concepts of the privacy ontology is represented by the ORM diagram of Figure 5.16.

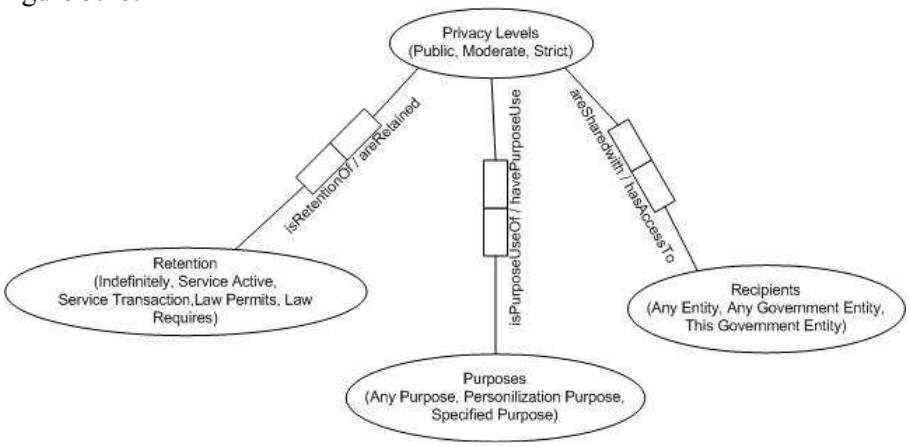


Figure 5.16 ORM diagram of the privacy ontology

Figure 5.16 shows the relation of privacy levels with “Recipients” (who is accessing the data), “Purposes” (what the data is accessed for) and “Retention” (how long the data are kept). With the previous definitions, the last step to build the privacy ontology is to specify the ontology in a platform like Protégé in order to generate the XML code.

Figure 5.17 shows the privacy ontology specified in Protégé platform; for example the highlighted concept in the left side window is the **Strict** privacy level. The right side window shows the relationships with other concepts such as use purpose, retention period and recipients who have access to the personal data.

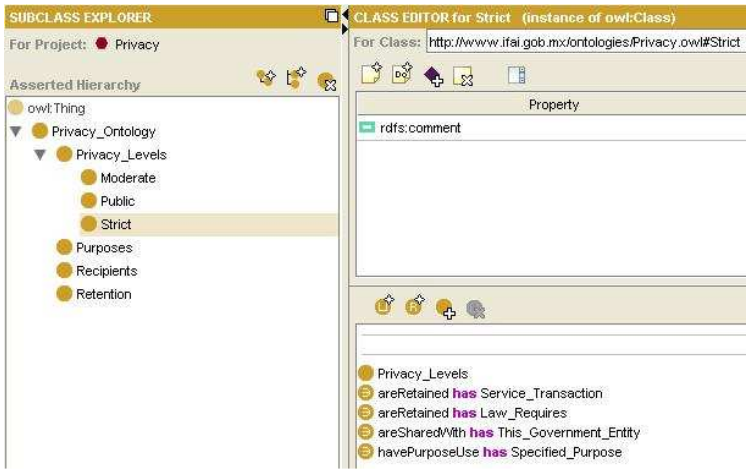


Figure 5.17 Privacy ontology specified in Protégé platform

Figure 5.18 shows the RDF/OWL code for the **Strict** privacy level.

RDF/XML Source Code

http://www.ifai.gob.mx/ontologies/Privacy.owl

Use XML Entities

Source Code

```
1: <rdf:RDF xmlns="http://www.ifai.gob.mx/ontologies/Privacy.owl#"

2:    <owl:Class rdf:ID="Strict">
3:      <rdfs:subClassOf rdf:resource="#Privacy_Levels"/>

4:            <owl:Restriction>
5:                <owl:onProperty rdf:resource="#havePurposeUse"/>
6:                <owl:hasValue rdf:resource="#Specified_Purpose"/>
7:            </owl:Restriction>
8:            <owl:Restriction>
9:                <owl:onProperty rdf:resource="#areSharedWith"/>
10:               <owl:hasValue rdf:resource="#This_Government_Entity"/>
11:           </owl:Restriction>
12:           <owl:Restriction>
13:               <owl:onProperty rdf:resource="#areRetained"/>
14:               <owl:hasValue rdf:resource="#Service_Transaction"/>
15:           </owl:Restriction>
16:           <owl:Restriction>
17:               <owl:onProperty rdf:resource="#areRetained"/>
18:               <owl:hasValue rdf:resource="#Law_Requires"/>
19:           </owl:Restriction>

20:   </owl:Class>

21: </rdf:RDF>
```

Figure 5.18 Fragment of the XML source code of the privacy ontology

Line 1 specifies the URI value for the ontology. Lines 2 and 3 define the privacy level **Strict** as a subclass of **Privacy_Levels**. Lines 4 to 7 specify the relationship with **Purposes** property indicating that attributes with this privacy level can be used only for the specified purpose for what the data were collected; the lexon 3.2.1 from Table 5.5 is expressed with this part of the code. Lines 8 to 11 describe the relationship with **Recipients** property pointing out that attributes with strict privacy level only can be shared with the government entity that is providing the service and that attributes cannot be shared with any other third party; the lexon 3.1.1 from Table 5.5 is expressed with this part of the code. Lines 13 to 19 specify the relationship with **Retention** property indicating that attribute can be retained only as long as the law requires or the service transaction lasts. This part of the code expresses in XML the lexons 3.3.3 and 3.3.4 of Table 5.5.

At this point, the phase I is completed where the personal information, legal framework and privacy ontologies are already expressed in a formal ontology language.

5.5.4 Phase II: Exchange of ontologies

The business agreement established during the building of the CoT specifies the ontology language and transfer protocol used to exchange ontologies. The IdP sends its personal information subset ontology to each SP with whom it is allowed to exchange attributes. In its answer, each SP sends back its services ontology specifying the services that are provided along with the personal data required to complete them.

As explained in section 5.5, the case scenario includes the entity government A (EgovA) playing the role of IdP and entity government B (EgovB) as the SP who demands attributes from EgovA. As an illustration, Figure 5.19 shows the personal information ontology specified by EgovA (it includes only the attributes handled by the entity) and the services ontology specified by EgovB.

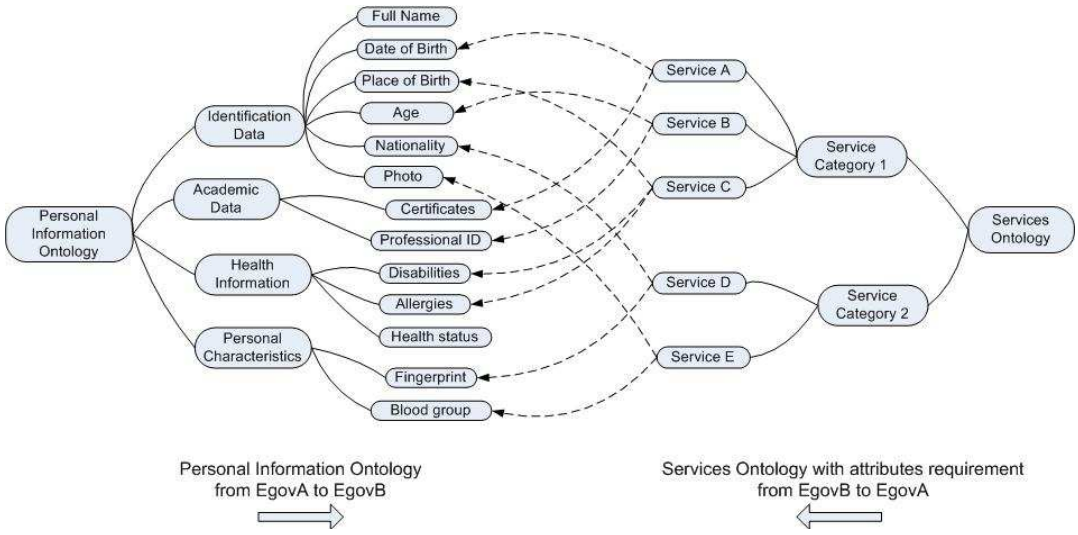


Figure 5.19 Exchange of PI and services ontologies

As shown in Figure 5.19, the EgovA (IdP) sends to EgovB (SP) its personal ontology (a subset of the PI ontology specified by IFAI). EgovB receives the PI ontology and relates its services ontology with the attributes required for each service; for example, the SP in order to provide *Service D* to the citizen requires *Nationality* and *Fingerprint* attributes. The services ontology is then sent to EgovA. This process is carried out initially, or each time the personal data structure changes or a service is modified or created. The completion of this phase allows the IdP to construct the privacy policy ontology presented in the next section.

5.5.5 Phase III: Generation of the privacy policy ontology

In this phase, the EgovA entity is able to build the default privacy policy ontology which relates the attributes of the citizen with the services that are allowed to access the attributes. From Figure 5.19 and the personal information described in Table 5.2, the Table 5.6 can be constructed showing the relationship between attributes, privacy levels, government entities and services resulted from the exchange of ontologies performed at phase II.

PI Category	Attribute	Privacy Level	Government Entity	Service
Identification Data	Full Name	Public	Any entity	Any
Identification Data	Date of Birth	Public	Any entity	Any
Identification Data	Place of Birth	Public	Any entity	Any
Identification Data	Age	Public	Any entity	Any
Identification Data	Nationality	Public	Any entity	Any
Identification Data	Photo	Public	Any entity	Any
Academic Data	Certificates	Moderate	Any Gov entity	Any
Academic Data	Professional ID	Moderate	Any Gov entity	Any
Health Information	Disabilities	Strict	EgovB	Service C
Health Information	Allergies	Strict	EgovB	Service C
Health Information	Health Status	Strict	None	None
Personal Characteristics	Fingerprint	Strict	EgovB	Service D
Personal Characteristics	Blood Group	Strict	EgovB	Service E

Table 5.6 Relationship between attributes and authorized access

In this case, *Public* data can be accessed by any entity in order to provide any service; *Moderate* data can be accessed only by government entities that provide any service; meanwhile, *Strict* data is accessed only by the government entity communicating its ontologies, in this example EgovB. For *Health Status* attribute, the access in this privacy policy ontology is denied as the privacy level is *Strict* and EgovB does not require the attribute. With the information contained in Table 5.6, the *Metadata Specification* module (see section 4.2.1 of chapter four) can build the privacy policy ontology as shown in Figure 5.20 in Protégé platform.

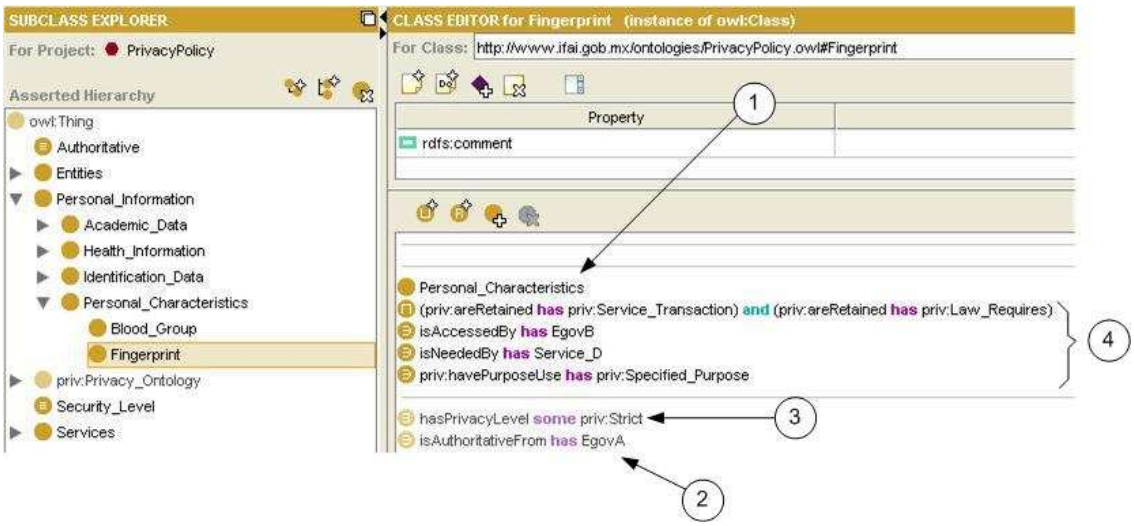


Figure 5.20 Privacy policy ontology presented in Protégé platform

The left side window of the Protégé screen shows a highlighted attribute (**Fingerprint**) whose relationships are expressed at the right side window. In this case, the attribute

Fingerprint is part of the **Personal_Characteristics** category (1); **EgovA** is the authoritative entity for the attribute playing the role of IdP (2); the attribute has a **Strict** privacy level (3); finally, the privacy parameters are shown indicating that only **Service_D** from **EgovB** entity can access the attribute, and that it is accessed for the specified purpose and that the retention time is only for the duration of the service transaction (**Service_Transaction**) or required by the law (**Law_Requires**).

The privacy policy ontology can be expressed in a standard XML ontology language as shown in Figure 5.21. This code is the base for building the privacy policy which is enforced each time the attribute is requested. This ontology can be modified at any time the citizen wants to change his privacy preferences; this process is explained in the next section. The XML code of Figure 5.21 corresponds to attribute Fingerprint of the privacy policy ontology.

RDF/XML Source Code

http://www.ifai.gob.mx/ontologies/PrivacyPolicy.owl

Use XML Entities

Source Code

```
1: <rdf:RDF xmlns="http://www.ifai.gob.mx/ontologies/PrivacyPolicy.owl#"

2:    <owl:Class rdf:ID="Fingerprint">
3:      <rdfs:subClassOf rdf:resource="#Personal_Characteristics"/>

4:            <owl:Restriction>
5:                <owl:onProperty rdf:resource="#isAccessedBy"/>
6:                <owl:hasValue rdf:resource="#EgovB"/>
7:            </owl:Restriction>
8:            <owl:Restriction>
9:                <owl:onProperty rdf:resource="#isNeededBy"/>
10:               <owl:hasValue rdf:resource="#Service_D"/>
11:           </owl:Restriction>
12:           <owl:Restriction>
13:               <owl:onProperty rdf:resource="&priv;havePurposeUse"/>
14:               <owl:hasValue rdf:resource="&priv;Specified_Purpose"/>
15:           </owl:Restriction>
16:           <owl:Class>
17:                   <owl:Restriction>
18:                       <owl:onProperty rdf:resource="&priv;areRetained"/>
19:                       <owl:hasValue rdf:resource="&priv;Law_Requires"/>
20:                   </owl:Restriction>
21:                   <owl:Restriction>
22:                       <owl:onProperty rdf:resource="&priv;areRetained"/>
23:                       <owl:hasValue rdf:resource="&priv;Service_Transaction"/>
24:                   </owl:Restriction>
25:           </owl:Class>
26:   </owl:Class>

27: </rdf:RDF>
```

Figure 5.21 XML code for the privacy policy ontology

Line 1 is the specification of the URI for the privacy policy ontology. Lines 2 and 3 describe the attribute **Fingerprint** as an element of the category **Personal_Characteristics**. Lines 4 to 11 indicate that the attribute is accessed only by **Service_D** of government entity **EgovB**. Lines 12 to 15 express the use purpose; in this case it is used only for the purpose specified by the service. Finally, lines 16 to 25 define the retention properties for the attribute.

This ontology represents the default privacy values for each attribute handled by EgovA regarding its relationship with EgovB. At this point, EgovA acting as IdP is able to allow the citizens to verify the privacy policies associated to their personal data; in addition, the citizens can express their privacy preferences changing the privacy level to some attributes as explained in the next section.

5.5.6 Phase IV: Citizen privacy preferences

The privacy policy ontology generated at phase III associates personal data with default privacy parameters regarding access, purpose and retention. This ontology is the basis for building the privacy policy used for the policy enforcement mechanism; this process is explained in next section.

Once the default privacy policy ontology is constructed, the citizens (previous authentication) can express their privacy preferences by modifying the privacy levels of some attributes. Figure 5.22 shows a possible graphical user interface (GUI) that allows citizens to express their privacy preferences, to input personal data or to verify the event log regarding their personal information usage. The module of the privacy model involved in this phase is the *Personal Information and Policy Access* module (see section 4.2.3 of chapter four).

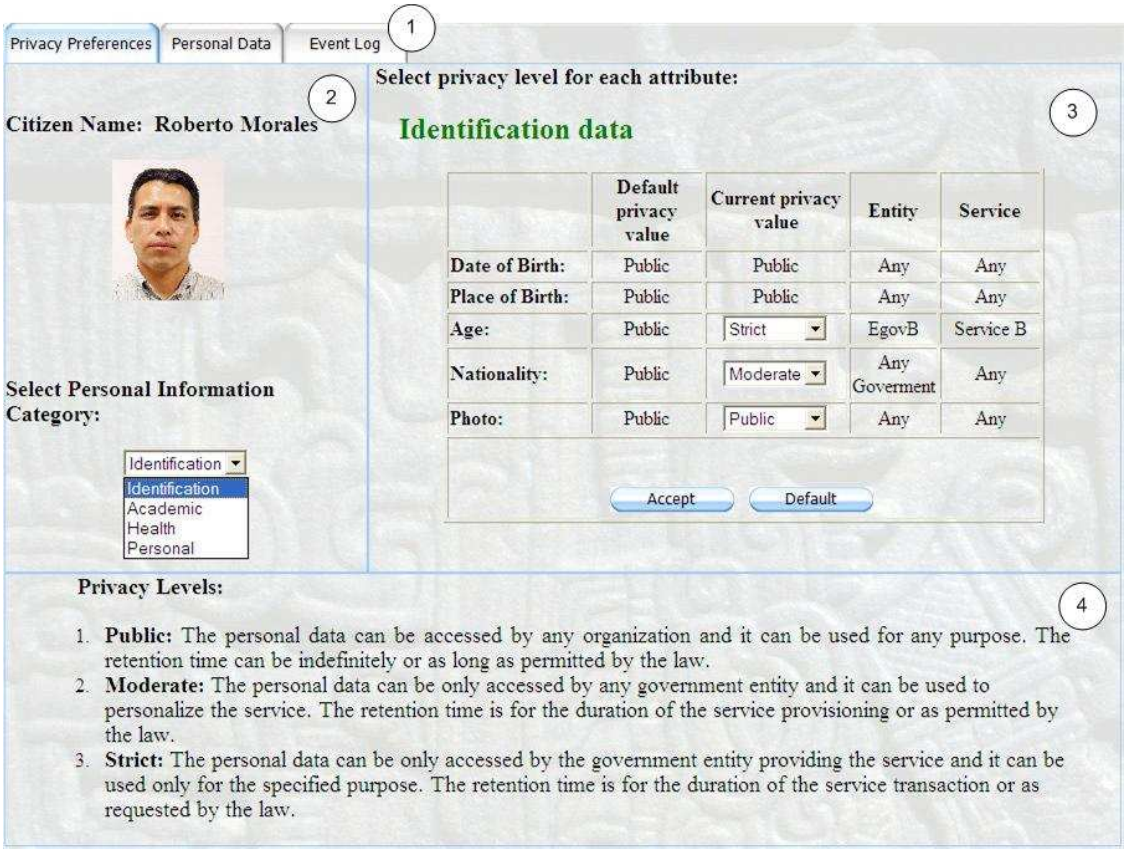


Figure 5.22 Graphical User Interface for expressing citizens´ privacy preferences

Figure 5.22 is a possible GUI provided by entity EgovA to allow citizens to express their privacy preferences. The tabs (1) indicate that this interface can also be used to input personal data or to access the logs, so that citizens can verify how their personal information is handled in terms of privacy. Window (2) shows the name and photograph of the citizen and the personal information category managed by EgovA; in this case, the citizen selected the **Identification** category in order to modify the privacy level of the attributes. The right upper window (3) presents the attributes composing the **Identification** category along with their privacy values. The default privacy value is the privacy level assigned initially by EgovA; in this example, all values are **Public** as defined by IFAI for this category. Next column represents the modified privacy value for each attribute; in this case the first two attributes are not allowed to be changed so they remain with **Public** value. For the remaining attributes, the citizen selected changing from **Public** to **Strict** for attribute **Age**, from **Public** to **Moderate** for attribute **Nationality** and attribute **Photo** was unchanged. The citizen can modify at any time those

values; when the **Accept** button is selected, the privacy preferences are stored and the privacy policy is generated. If the **Default** button is selected the default privacy values are reassigned to the attributes. The last two columns show the entity and the service that are allowed to access the attributes depending on the associated privacy level.

The bottom window (4) explains in natural language the privacy parameters for each level so that citizens can select the appropriate values according to their privacy preferences.

5.5.7 Phase V: Generation of the privacy policies

After the citizens have optionally specified their privacy preferences at phase IV, the privacy policy ontology is mapped to a privacy policy expressed in an XML access control language ready to be used with the enforcement mechanism. The *Privacy Policy Management* module is responsible for making this policy mapping (see section 4.2.2 of chapter four).

5.5.7.1 Mapping from privacy policy ontology to privacy policy

As described in section 3.3.3 of chapter three, a privacy policy has a structure with four sections: the *Combining Algorithm*, the *Target*, the set of *Rules* and *Obligations*; meanwhile, the privacy policy ontology contains a set of elements definitions and relationships with privacy parameters. The mapping can be carried out as follows:

1. The *Combining Algorithm* determines the way the results of each rule are combined in order to get a unique response; it can be a fixed value for all the policies. For example, the value can be *Deny Overrides* which establishes that if any rule is evaluated to deny, the combined result is deny; this is a general criteria where all the rules must be satisfied.
2. The *Target* has three subsections: *Resource* corresponds to the attribute to be accessed; *Subject* relates with the *Recipients* (entities and services allowed to access the attributes); and *Action* indicates the operation to be performed with the attribute.
3. The *Rules* are filled out with *Purpose* and *Retention* privacy conditions from the ontology.
4. The *Obligations* section can be fixed to the log event action in order to fulfill with log requirements of the privacy model, as defined by the *Personal Information and Policy Log module* (see section 4.2.4 of chapter four)

Figure **5.23** shows graphically the mapping between the privacy policy ontology elements and the privacy policy sections:

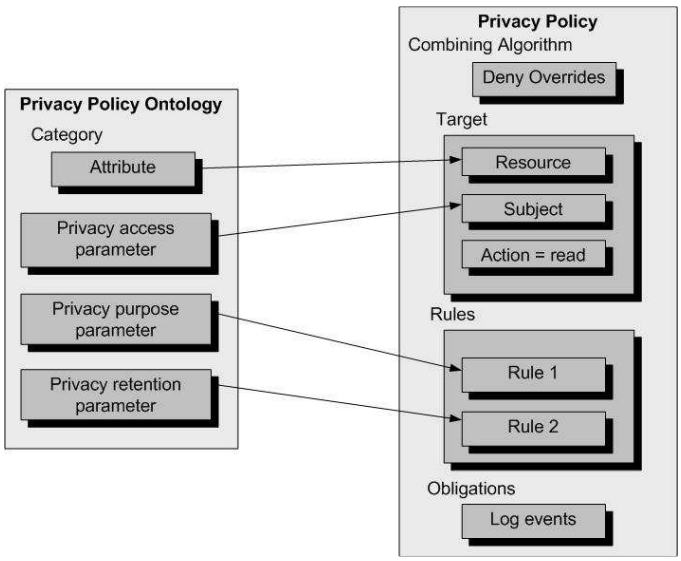


Figure 5.23 Mapping from privacy policy ontology to privacy policy

As an example, Figure 5.24 shows the mapping from the fragment ontology of Figure 5.21 to a privacy policy expressed in an XML access control language such as XACML (Moses, 2005).

Combining Algorithm

```
[1]  <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
[2] <Policy PolicyId="example1:PersonalDataPolicy"
RuleCombiningAlgId="urn:oasis:names:tc:1.0:rule-combining-algorithm:deny-overrides">
[3]  <Description>This policy defines the privacy parameters for the attribute Fingerprint of the
Personal_Characteristics category.</Description>
```

Target

```
[4]  <Target>

[5] <Resource>
[6] <ResourceMatch MatchId="urn:oasis:names:tc:2.0:function:anyURI-regexp-match">
[7]   <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">
Personal_Characteristics.Fingerprint</AttributeValue>
[8]   </ResourceMatch>
[9] </Resource>

[10] <Subject>
[11] <SubjectMatch MatchId="urn:oasis:names:tc:1.0:function:string-equal">
[12]  <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">
EgovB.Service_D</AttributeValue>
[13]  </SubjectMatch>
[14]  </Subject>

[15] <Action>
[16] <ActionMatch MatchId="urn:oasis:names:tc:1.0:function:string-equal">
[17]  <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">
read</AttributeValue>
[18]  </ActionMatch>
[19]  </Action>

[20]  </Target>
```

Rules

```
[21] <Rule Effect="Permit" RuleId="urn:oasis:names:tc:2.0:matching-purpose">
[22] <Condition>
[23] <Apply FunctionId="urn:oasis:names:tc:1.0:function:equal">
[24]  <ActionAttributeDesignator DataType="http://www.w3.org/2001/XMLSchema#string"
AttributeId="urn:oasis:names:tc:2.0:action:purpose" />
[25]  </Apply>
[26]  <AttributeValue
DataType="http://www.w3.org/2001/XMLSchema#string">Specified_Purpose</AttributeValue>
[27]  </Condition>
[28]  </Rule>

[29] <Rule Effect="Permit" RuleId="urn:oasis:names:tc:2.0:matching-retention">
[30] <Condition>
[31] <Apply FunctionId="urn:oasis:names:tc:1.0:function:equal">
```

```
[32]  <ActionAttributeDesignator DataType="http://www.w3.org/2001/XMLSchema#string"
AttributeId="urn:oasis:names:tc:2.0:action:retention" />
[33]  </Apply>
[34]  <AttributeValue
DataType="http://www.w3.org/2001/XMLSchema#string">Law_Requires</AttributeValue>
[35]  </Condition>
[36]  </Rule>

[37] <Rule Effect="Permit" RuleId="urn:oasis:names:tc:2.0:matching-retention">
[38] <Condition>
[39] <Apply FunctionId="urn:oasis:names:tc:1.0:function:equal">
[40]  <ActionAttributeDesignator DataType="http://www.w3.org/2001/XMLSchema#string"
AttributeId="urn:oasis:names:tc:2.0:action:retention" />
[41]  </Apply>
[42]  <AttributeValue
DataType="http://www.w3.org/2001/XMLSchema#string">Service_Transaction</AttributeValue>
[43]  </Condition>
[44]  </Rule>
```

Obligations

```
[45] <Obligations>
[46] <Obligation ObligationId="example1:obligation:log" FulfillOn="Permit">
[47] <AttributeAssignment AttributeId="example1:attribute:log"
DataType="http://www.w3.org/2001/XMLSchema#string">
[48]  <AttributeAssignment AttributeId="example1:attribute:text"
DataType="http://www.w3.org/2001/XMLSchema#string">Your Attribute has been accessed
by:</AttributeAssignment>
[49]  <SubjectAttributeDesignator DataType="http://www.w3.org/2001/XMLSchema#string"
AttributeId="urn:oasis:names:tc:1.0:resource:subject-id" /> </AttributeAssignment>
[50]  </Obligation>
[51]  </Obligations>
```

Figure 5.24 Privacy policy expressed in XACML language

Lines [1] to [3] identify the policy and define the combining algorithm as “deny-overrides”. Lines [4] to [20] describe the *Target* section, where the *Resource* subsection corresponds to the **Fingerprint** attribute of the **Personal_Characteristics** category (lines [5] to [9]); meanwhile lines [10] to [14] identify the entity and service that are allowed to access the attribute (**Service_D** from **EgovB**); lines [15] to [19] define that **read** is the default action for all the privacy policies.

Lines [21] to [44] represent the rules that are evaluated for the *Purpose* and *Retention* privacy conditions. Finally, lines [45] to [51] express the obligations that must be fulfilled when the policy applies and it is evaluated to true, in this case the obligation is that each transaction must be logged.

Once the privacy policy is constructed, it is ready to be used by the enforcement mechanism. Section 3.3.3 of chapter three explains how the request parameters must match the *Resource*, *Subject* and *Action* values so that the policy can be applied and the rules evaluated.

5.6 Conclusions

The case scenario presented in this chapter showed how the federated identity architecture can simplify the management of digital identities within a collaborative environment such as the eGovernment services. In addition, the privacy model proposed in chapter four can be applied for improving the privacy when personal data are exchanged between different government entities.

Mexican government has started some efforts in the legal and technological fields in order to guarantee the privacy to some extent when personal information of the citizens is handled by the federal public administration. However, there is no real integration between the privacy legal framework and the technological mechanisms currently implemented.

The methodology presented at the last part of the chapter, shows how to develop the privacy model for the e-Government services deployment. The process is divided into five phases: the first phase describes how the privacy requirements generally expressed in natural language can be converted to a formal expression using ontologies; the second phase presents the ontology exchanged for matching the services provided by the SPs and the personal data required from the IdPs in order to complete them; the third phase shows how the ontologies can be combined for building the privacy policy ontology which is the basis for the final privacy policy; the fourth phase proposes a possible graphical user interface so that citizens are able to express easily their privacy preferences; finally, the last phase demonstrates how the privacy policy ontology can be mapped to a privacy policy that can be enforced.

The methodology also shows the use of diverse standard languages and tools such as RDF, OWL, ORM, Protégé platform and XACML. The privacy model is implemented in a Mexican e-Government context; however, it can apply to any environment where privacy of personal data is an important requirement.

5.7 Conclusions (en français)

Le cas d'étude présenté dans ce chapitre a montré comment l'architecture d'identité fédérée peut simplifier la gestion des identités numériques dans un environnement de collaboration tel que les services d'e-Gouvernement. De plus, le modèle de respect de la vie privée proposé au chapitre quatre peut être appliqué pour améliorer le respect de la vie privée quand les données personnelles sont échangées entre différentes entités gouvernementales.

Le gouvernement mexicain a commencé quelques efforts dans les domaines légaux et technologiques afin de garantir le respect de la vie privée quand les informations personnelles des citoyens sont traitées par l'administration publique fédérale. Cependant, il n'y a pas une vraie intégration entre le cadre juridique de respect de la vie privée et les mécanismes technologiques actuellement mis en application.

La méthodologie présentée à la dernière partie du chapitre, montre comment développer le modèle de respect de la vie privée pour le déploiement de services d'e-gouvernement. Le processus est divisé en cinq phases : la première phase décrit comment les exigences de respect de la vie privée généralement exprimées en langage naturel peuvent être exprimées formellement grâce aux ontologies ; la deuxième phase présente l'échange d'ontologies entre les SPs qui fournissent une ontologie de leurs services et les IdPs qui fournissent les données personnelles requises par les SPs pour mener à bien le service; la troisième phase montre comment les ontologies peuvent être combinées pour établir l'ontologie de la politique de protection de la vie privée dans le cas d’une transaction précise, ce qui permet de construire la politique de protection de la vie privée finale; la quatrième phase propose une interface utilisateur graphique de telle sorte que les citoyens puissent exprimer facilement leurs préférences quant au respect de la vie privée ; finalement, la dernière phase démontre comment l'ontologie de politique de respect de la vie privée peut être dérivée en une politique XACML de vie privée prête à être appliquée.

La méthodologie montre également l'utilisation des langages et des outils standards tels que RDF, OWL, ORM, plate-forme de Protégé et XACML. Le modèle de respect de la vie privée a été mis en application dans un contexte mexicain d'e-gouvernement. Cependant, il peut être appliqué à n'importe quel environnement où le respect de la vie privée des données personnelles est une exigence forte.

Chapter 6

Conclusions

The existence of multiple digital identities for a unique user in a distributed and collaborative environment such as Internet represents a real drawback for users and Service Providers. This problem has limited in some extent a faster development of online services.

It is required that an Identity Management System capable of managing multiple identities simplifying their management and use for users and SPs. Among the IMS models proposed, the Federated Identity Architecture and the User-Centric models fulfill the actual online and collaborative requirements for distributed identity management. The FIA focuses its function on the composition of circles of trust through the use of business agreements and a common technological platform. The specifications of the FIA allow the entities of the CoT to make use of identity services such as single sign on, identity federation and attribute sharing. The FIA operation is targeted to facilitate the interaction between organizations composing the FIA. Therefore users have little participation mainly at the privacy specifications regarding how their personal data must be handled. FIA architecture is based on standards simplifying the identity management and enhancing the user experience due to the identity services provided; however, it limits the scalability due to the restricted business agreements among the entities. On the other hand, User-Centric model centralizes its function on the user which must approve each piece of personal data that is transferred from one entity to another. This model scales very well and ensures a high level of privacy, but the most important drawback is that users must be online in all transactions and all the security issues fall in the user device.

Due to its characteristics and function, the FIA is considered as the best solution for managing digital identities and for deploying identity services in a distributed environment where entities must share personal information in a secure way. Among all the FIA initiatives, Liberty Alliance is the most solid specification. It is based on open standards allowing that commercial solutions as well as open source products can interoperate in a heterogeneous environment. Liberty Alliance is a mature specification supported by many vendors from different areas and it is oriented to enterprise environment. Shibboleth is another federated initiative focused to academic purposes that simplifies the information exchanged mainly for academic collaboration. This initiative does not take care of the real identity of the user to provide him a service; rather it relies on the group membership of the user. The last proposal made by Microsoft complements the stack of Web Services specification for managing digital identities in a federated architecture. All the initiatives are based on open standards so that the interoperability is guaranteed. However, Liberty Alliance is the most implemented specification in available products and it has been deployed in many projects involving federated architectures.

As it was shown in chapter two, the federated architectures simplify the personal information (attributes) sharing among the entities composing the CoT. However, it can represent a serious risk in terms of privacy if there are not sufficient mechanisms that guarantee the security of such personal information.

Nowadays, privacy is a very important issue for users in order to consolidate the online services development. FIA has inherent security mechanisms (or Privacy Enhancing Technologies) that improve the privacy level of the infrastructure; however, they are not sufficient if the privacy requirements specified in a privacy legal framework must be fulfilled.

Regarding privacy regulatory frameworks, since last century, two legal models have been developed. The first legal model supported by Europe is a general model that applies to private and public sectors and it proposes universal privacy principles. In this model, the governments play an important role supervising the fulfillment of the regulatory framework. The other legal model supported mainly by United States is a sectorial approach with auto-regulation function. This model is specific to each sector so that governments play a partial role in its supervision. Both models have influenced the development of regulatory frameworks all around the world; Mexico has followed the European model with some modifications pretending to adopt a hybrid model.

Independently of the type of privacy legal framework that must be fulfilled in a particular case, the main challenge is that a technological infrastructure must support and enforce the compliance of the privacy requirements dictated by the legal framework.

The main contribution of the present work is the proposition of a model based on privacy policies for enhancing the privacy within a federated identity architecture targeted to manage digital identities in a distributed and collaborative environment. The proposed model allows that a technological infrastructure is compliant to a privacy legal framework. The model is an architecture composed of layers and modules with function specifications based on open standards. The main objective of the model is that personal information sharing within a FIA is protected by privacy policies. That is, when an attribute is requested from one entity to another, the policy is evaluated for verifying the privacy compliance before releasing the attribute.

The model can be implemented centralized with the components of the IdP or it can be distributed across all entities composing the CoT. This modularity facilitates its implementation and scalability; in conclusion, it is a flexible architecture that can integrate any regulatory framework.

Among the benefits provided by the privacy model for enhancing the privacy in a Federated Identity Architecture, there could be mentioned the following ones:

- It allows translating regulatory frameworks to a formal expression such as ontology. With the methodology used, any regulatory framework can be translated giving enough flexibility to support privacy requirements.
- Since it is based on privacy policies, it protects the personal information by enforcing the policies each time an attribute is requested. Such privacy policies are created with the specification of regulatory frameworks, privacy requirements of the organizations and the privacy preferences of the user
- The model is centralized on users allowing them to express their privacy preferences through the use of an easy graphical interface. Additionally, the users can consult their personal information and the privacy policies associated, so they know at any time how their personal information is handled regarding privacy.
- The model is a layered architecture with well specified functions for each module, this facility permits deploying the modules in a centralized way; for instance at the IdP component or they can be distributed across the entities of the CoT.
- The specification of the function of each module is platform independent so they can be implemented with solutions based on open standards such as those used for the development of web services. This guarantees the interoperability of solutions built with some heterogeneous technology.

- The model specifies that all events generated by its components must be logged, so that users and auditors can access the event log to verify the compliance with the established privacy policies.

The modules of the lowest layer translate the privacy requirements (expressed in natural language from regulatory entities, organizations and user´s privacy preferences) to a formal expression represented by an ontology standard language. The ontologies are exchanged and combined in order to correlate the personal information with privacy requirements and services accessing the attributes. The resulting ontology expresses what attributes may be accessed by what service and under what privacy parameters. This privacy ontology is taken as a base for building the privacy policy associated to the corresponding personal information. At this point, users can verify what policies are applied to their personal data and therefore how they are handled in term of privacy. Additionally, users can modify some privacy parameters expressing in this way their privacy preferences. When the privacy policies are optionally modified by users, they are ready to protect their personal data each time an attribute is requested.

The policy creation and enforcement are carried out by the modules of the second layer of the model. The privacy policies can be modified at any time when users change their privacy preferences or the regulatory requirements are updated. This means that the privacy model must support a dynamic environment.

The upper layer of the model provides an interface that allows users to interact with the architecture in order to modify their personal data, to verify the privacy policies associated to their personal information, to express their privacy preferences and to consult the event logs so they know how their personal data are treated in terms of privacy.

The case scenario described in chapter five showed how to implement the privacy model within a federated architecture for Mexican e-government service deployment. However, the privacy model can be applied to any case scenario. The personal data profile was taken from the personal information specifications made by IFAI. The privacy levels associated to personal categories were assigned according to the security levels defined for each category. In this case, three privacy levels were specified indicating different privacy parameters for describing data for public access, access from any government entity or access from only specific government entity. The number of privacy levels and the privacy parameters for each level can be defined according to the specific case scenario, for example the health sector, financial sector, academic sector, etc.

A methodology developed by a European organization was used to build ontologies from specifications expressed in natural language. This is an important issue so that any regulatory framework can be expressed in formal ontology language. The ontologies were combined and the resulting policy ontology was translated to a privacy policy expressed in a policy language able to be evaluated and enforced. A graphical user interface was proposed for allowing citizens to select a personal data category and modify the privacy requirements assigned by default.

The development of the case scenario was very important to demonstrate the feasibility of implementing the model into a federated architecture. In conclusion, the privacy model developed in this work is ready to be deployed in any case scenario where a federated identity architecture is used as an identity management system. The model can be applied within the Mexican e-Government scenario where online services are deployed for citizens in a collaborative and distributed environment.

Limitations and future works

- It is necessary to define and evaluate the performance parameters. In an environment with high degree of transactions requesting attributes, the process of evaluating and enforcing privacy policies at the IdP may constitute a bottleneck point with the corresponding degradation of time response.
- The integration of new modules and functions within a FIA may introduce security vulnerabilities. Therefore, it is very important to assure that all the communications between the elements of the model are secure, that is, that their communications are

authenticated and encrypted. The graphical interface that allows users and auditors to access personal information, privacy policies and event logs must incorporate robust authentication mechanisms for preventing that unauthorized people can modify the personal information or change its privacy level.

- The privacy model is targeted for a dynamic environment where regulatory frameworks, privacy policies and personal data change constantly. Therefore, it is necessary to guarantee the consistency of policies and personal data within all the elements of the CoT. The problem is more serious when the personal data are interchanged between two or more different CoT.
- The auditing process is limited to the online interaction of users and auditors with the components of the model; however, it is possible to integrate automatic auditing process by adding communication interfaces to the log module.
- The model could cope with scalability issues when it is extended for supporting inter CoTs interactions due to the complexity for constructing trust relationships throughout all the components.
- Despite that the specification of the model is based on open standards, there might be interoperability problems due to the different platforms used for deploying federated architectures.

Publications

[1] U. Fragoso-Rodriguez, M. Laurent-Maknavicius, J. Incera-Dieguez, "Federated Identity Architectures", Proceedings Mexican Conference on Informatics Security, Mexico DF, November 14-17, 2006.

[2] U. Fragoso-Rodriguez, M. Laurent-Maknavicius, J. Incera-Dieguez, "Federated Identity Architectures Evaluation", Works in Progress Session, Annual Computer Security Applications Conference, Miami Florida, December 11-15, 2006.

[3] U.Fragoso-Rodriguez, “A survey on privacy of personal information and its implications in federated identity architectures”, Proceedings of 9th Annual Global Information Technology Management Association (GITMA) World Conference 2008, Atlanta Georgia USA, June 22-24, 2008.

[4] U. Fragoso-Rodriguez, M. Laurent-Maknavicius, J. Incera-Dieguez, “Privacy Management Model in Federated Identity Architecture”, Works in Progress Poster Session, 50 years of computing in Mexico and 25 years of computing in the CINVESTAV, Mexico DF, September 1-5, 2008.

Under review for the Journal of Information Privacy and Security (JIPS):

[5] U.Fragoso-Rodriguez, “Privacy management model for Federated Identity Architectures”, October, 2009.

References

Agrawal, R., & al. (2003). *An XPath-based Preference Language for P3P*. WWW2003. May 20-24, 2003. Budapest Hungary.

Anderson, A. (2004). *An Introduction to the Web Services Policy Language (WSPL)*. Fifth IEEE International Workshop on Policies for Distributed Systems and Networks. June 07-09, 2004. Yorktown Heights, New York. (pp. 189).

Anderson2, A. (2005). *SAML 2.0 profile of XACML v2.0.* OASIS Standard. February 2005.

Andrews, S. (2003). *International Corporate Privacy Handbook.* Annual meeting of the American Bar Association. August 8-11, 2003. San Francisco, California.

Bhargav-Apantzel, A., & al. (2005). *Establishing and Protecting Digital Identity in Federation Systems.* The First ACM Workshop on Digital Identity Management -- DIM 2005. (pp. 269-300).

Candia, T. (2004). *Benefits of Federated Identity to Government.* Liberty Alliance Project. March 7, 2004.

Cantor, S. (2005). *Shibboleth Architecture, Protocols and Profiles.* Working draft 02. September 10, 2005. Available at: http://shibboleth.internet2.edu/shibboleth-documents.html

Chappell, D. (2006). *Introducing Windows CardSpace.* Microsoft Technical Articles. April 2006.

Commerce, D. (1998). *Safe Harbor Overview.* Retrieved from Department of Commerce U.S.: http://www.export.gov/safeharbor/SH_Overview.asp

Council, E. (1995). *Protection of individuals with regard to the processing of personal data and on the free movement of such data.* October 1995. Avilable at: http://eur-lex.europa.eu/

Cranor, L. (2002). *The Platform for Privacy Preferences 1.0 Specifications (P3P).* W3C working Draft. April 16, 2002.

Cranor2. (2002). *A P3P Preference Exchange Language 1.0 (APPEL 1.0).* W3C Working Draft. April 15, 2002.

Cuppens, F. (2006). *Les modèles de sécurité.* Sécurité des réseaux et systèmes répartis. Editions Hermes. MÉL.-DESWARTE Y.

Dongwan, S., & Gail-Joon, A. (2004). *Ensuring Information Assurance in Federated Identity Management.* IEEE International Conference on Performance, Computing and Communications, (pp. 821-826).

Dumortier, J., & Goemans, C. (2004). *Legal Challenger for Privacy Protection and Identity Management*. Proceedings of the NATO/NASTEC workshop on Advanced Security.

Duran, A. (2003). *How the Nature of Identity will Shape Its Deployment*. Digital ID World Conference. October 15-17, 2003. Denver, Co.

Eclipse. (2008). *Eclipse Releases Its First User-Centric Identity Framework*. February 2008. Available at: http://www.eclipse.org/org/press-release/20080221_higgins.php

eMexico. (2009). *El Sistema Nacional eMexico*. Retrieved from eMexico Portal: http://www.e-gobierno.gob.mx/wb2/eMex/eMex_El_Sistema_Nacional_eMexico_

Europe, C. (1981). *Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data.* January 1981. Strasbourg. Available at: http://conventions.coe.int/treaty/EN/Treaties/Html/108.htm

Federal, T. (1974). *Privacy Act.* Retrieved from Federal Trade Commission: http://www.ftc.gov/foia/privacy_act.html. December 1974.

Garfinkel, S., & Faith, L. (2002). *P3P: Privacy Primer.* Available at: http://www.oreillynet.com/network/excerpt/p3p/p3p.html

Guerrier, C. (2008). *Protection des données personnelles et environnement réseau*. Techniques de l´Ingénieur, Sécurité des systèmes d´Information. October 2008. Available at: http://www.techniques-ingenieur.fr/book/h5445/protection-des-donnees-personnelles-et-environnement-reseau.html

Health, D. (1996). *Summary of the HIPPA Privacy Rule.* Retrieved from Department of Health and Human services, U.S.: http://www.hhs.gov/ocr/privacysummary.pdf

Higgins. (2008). *Higgins overview (Power Point presentation).* Retrieved 2008, from Eclipse: http://www.eclipse.org/higgins/documents/Higgins-Overview-2008.ppt

IBM. (2002). *Security in a Web Services World: A Proposed Architecture and Roadmap.* IBM and Microsoft white paper. April 1, 2002.

IFAI. (2005). *Personal Data Protection Directives.* Mexican Official Gazette. September 30, 2005. (pp. 55-64).

IFAI2. (2003). *Guidelines for Security Personal Data Protection.* Instituto Federal de Acceso a la Información. Available at: http://www.ifai.org.mx/

IFAI3. (2002). *Federal Law of Transparency and Access to Government Public Information.* Instituto Federal de Acceso a la Información. June 11, 2002. Available at: http://www.ifai.org.mx/transparencia/LFTAIPG.pdf

Jean, L. (2004). *Digital Identity*. IEEE Technology and society Magazine . Vol. 23, No 3. (pp. 34-41).

Jonathan, T. (2004). *Liberty ID-WSF Web services Framework Overview.* Version 1.0. Liberty Alliance Project. Available at: http://www.projectliberty.org/liberty/content/download/

Josang, A. (2005). *Trust Requirements in Identity Management*. Australasian Information Security Workshop 2005. Newcastle, Australia. (pp. 99-108).

Kaler, C., & Nadalin, A. (2003). *Web Services Federation Language (WS-Federation).* Version 1.0. July 8, 2003.

Kumararugu, P., & al. (2007). *A survey of Privacy Policy Languages*. Workshop on Usable Privacy and Security. JUly 18-20, 2007. Carnegie Mellon University, Pittsburgh, PA.

Landau, S. (2003). *Liberty ID-WSF Security and Privacy Overview.* Liberty Alliance Project. Available at: http://www.projectliberty.org/liberty/files/specs/

Landesberg, M., & al. (1998). *Privacy Online: A report to Congress.* U.S: Federal Trade Commission. June 1998. Available at: http://www.ftc.gov/reports/privacy3/toc.shtm

Liberty, A. (2009). *E-Government*. Available at:

http://www.projectliberty.org/liberty/adoption/egovernment

Linn, J. (2004). *Trust Models Guidelines 2.0.* OASIS. February, 2004. Available at: http://www.oasis-open.org/committees/download.php/6158/sstc-saml-trustmodels-2.0-draft-01.pdf

Lippmann, G. (2005). *Interoperability of eGovernment systems*. Survey for the 44th meeting of the Directors general responsible for Public Administration of EU member states. June 10, 2005. Luxembourg.

Moses, T. (2005). *eXtensible Access Control Markup Language (XACML).* Version 2.0 OASIS. February 1, 2005. Available at: http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

OECD. (1980). *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.* September 23, 1980. Available at: http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html

OECD2. (2003). *Policy and Practical Guidance for Implementing Privacy Protection Online.* Privacy Online OECD Guidance on policy and practice. Available at: http://www.oecd.org/document/49/0,3343,en_2649_34255_19216241_1_1_1_1,00.html

Pato, J. (2003). *Identity Management: Setting Context*. Encyclopedia of Information Security. April 8, 2003.

Persona. (2006). *Persona Information System*. Available at: http://persona.ifai.org.mx/persona/welcome.do

Pope, S., & Josang, A. (2005). *User Centric Identity Management*. AusCERT Conference 2005, (pp. 1-3).

PRIME. (2009). *PRIME Overview*. Available at: http://www.prime-eu.org/index.html

Privacy, I. (2006). *PHR2005-Overview of Privacy.* October 10, 2006. Available at: http://www.privacyinternational.org/article.shtml

Protégé. (2009). *Protégé Overview*. Available at: http://protege.stanford.edu/overview/

Ragouzis, N. (2006). *SAML V2.0 Technical Overview.* OASIS Working Draft 10. October 9, 2006.

Scavo, T., & Cantor, S. (2005). *Shibboleth Architecture, Technical Overview.* Working Draft 02. June 8, 2005. Available at: http://shibboleth.internet2.edu/docs/draft-mace-shibboleth-tech-overview-latest.pdf

Segob. (2006). *Normative Guideline for CURP assignment.* Available at: http://www.renapo.gob.mx/pdf/InstructivoParaLaCurp.pdf

Senate. (2008). *Senate Gazzete No 308.* Available at: http://www.senado.gob.mx/gace.php?sesion=2008/12/04/1&documento=71

Senicar, V., & al. (2003). *Privacy-Enhancing Technologies approaches and development*. Computer Standards & Interfaces 25. May 2003. (pp.147-158).

Sobel, D. (2006). *The Federal Institute for Access To Information in Mexico and a Culture of Transparency.* A Report for the Williams and Flora Hewlett Foundation. February, 2006.

Subenthiran, S., Sandrasegaran, K., & Shalak, R. (2004). *Requirements for Identity Management in Next Generation Networks*. The 6th International Conference on Advanced Communication Technology, (pp. 138-142).

Sun, C. (2003). *The European Union Privacy Directive and Its Impact on the U.S. Privacy Protection Policy: A year 2003 Perspective*. Northwestern Journal of Technology and Intellectual Property , (pp. 99-116).

Tang, Y. (2005). *Ontology Development Process.* Privacy and Identity Management for Europe (PRIME).

Varney, C. (2003). *Privacy and Security Best Practices.* Version 2.0. Liberty Alliance Project. November 12, 2003.

Watson, T. (2003). *Introduction to the Liberty Alliance Identity Architecture.* Revision 1.0 Liberty Alliance Project. March, 2003. Available at: http://xml.coverpages.org/LibertyAllianceArchitecture200303.pdf

Windley, P. (2005). *Digital Identity* (First ed.). O´Reilly. ISBN: 0-596-00878-3. August, 2005.

Wordnet. (2008). *Lexical database for the English language.* Available at: http://wordnet.princeton.edu

List of acronyms

AAA	Authentication Authorization Auditing
ACM	Access Control Model
AP	Attribute Provider
APPEL	A P3P Preference Exchange Language
AR	Attribute Requester
ARP	Attribute Release Policies
B2B	Business to Business
B2C	Business to Client
B2E	Business to Employee
BB	Business Broker
CDM	Context Data Model
CoT	Circle of trust
CURP	Clave Unica de Registro de Información
DCC	Digital Community Centers
DI-WS	Data Interface Web Service
DM-WS	Data Mapping Web Service
FIA	Federated Identity Architecture
GI	Graphical Interface
GPS	Global Positioning Satellite
GUI	Graphical User Interface
HIPPA	Health Insurance Portability and Accountability Act
HTTP	Hyper Text Transfer Protocol
I-BAC	Identity Based Access Control
I-Card	Identity Card
ICT	Information and Communication Technology
ID-FF	IDentity Federated Framework
IdAS	Identity Attribute Service
IdP	Identity Provider
IdS	Identity Service
ID-WSF	IDentity Web Service Framework
ID-SIS	IDentity Services Interface Specifications
IETF	Internet Engineering Task Force
IFAI	Instituto Federal de Acceso a la Información
IMS	Identity Management System
IPsec	Internet Protocol security
LC-WS	Log Classification Web Service
LFO	Legal Frame Ontology
LI-WS	Log Interface Web Service
LS-WS	Log Store web Service
MFPA	Mexican Federal Public Administration
MS	Metadata Specification
OASIS	Organization for the Advancement of Structured Information Standards
OECD	Organization for Economics Co-operation and Development
OE-WS	Ontology Exchange Web Service
OM-WS	Ontology Merging Web Service
Or-BAC	Organizational Based Access Control
ORM	Object Role Modeling
OWL	Ontology web Language
P2P	Peer To Peer
P3P	Privacy Preference Project

PAP	Policy Administration Point
PDP	Policy Decision Point
PDP-WS	Policy Decision Point web Service
PDS	Personal Data Systems
PEL	Privacy Expression Language
PEP	Policy Enforcement Point
PEP-WS	Policy Enforcement Point web Service
PET	Privacy Enhancing Technologies
PG-WS	Policy Generator web Service
PI	Personal Information
PIDA	Personal Information Data Abstraction
PIM	Personal Information Management
PIN	Personal Identifier Number
PIP	Policy Information Point
PIPA	Personal Information and Policy Access
PIPL	Personal Information and Policy Log
PIPO	Personal Information Profile Ontology
PIT	Privacy Intrusion Technologies
PI-WS	Personal Information Web Service
PKI	Public Key Infrastructure
PO	Privacy Officer
POI-WS	Policy Interface Web Service
PPE	Privacy Policy Enforcement
PPL	Privacy preference Language
PPM	Privacy Policy Management
PPO	Privacy Policy Ontology
PPS	Privacy Policy Set
PRIME	PRivacy and Identity Management for Europe
PRO	Privacy Ontology
PSI-WS	Policy Storing and Indexing Web Service
R-BAC	Role Based Access Control
RDF	Resource Description Framework
RENAPO	Registro Nacional de Población
RFID	Radio Frequency ID
RP	Relying Party
SAML	Security Assertion Markup Language
SIN	Single Identification Number
SLO	Single Log Out
SOA	Service Oriented Architecture
SOAP	Simple Object Access Protocol
SP	Service Provider
SPO	Service Profile Ontology
SSL	Secure Socket Layer
SSN	Social Security Number
SSO	Single Sign On
STS	Security Token Server
TLS	Transport Layer Security
UIC	Unique Identity Card
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
V-BAC	View Based Access Control
W3C	World Wide Web Consortium
WAYF	Where Are You From
WSC	Web Service Consumer
WSP	Web Service Provider

WSPL	Web Service Policy Language
XACML	eXtension Access Control Markup Language
XML	eXtension Markup Language
XPref	X-Path Preference Language

List of figures

List of tables