UNIVERSIDAD POLITÉCNICA DE MADRID

**ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN**

**ARQUITECTURA Y MODELO DE REFERENCIA PARA
UN ECOSISTEMA DE INTERCAMBIO DE DATOS
PERSONALES CENTRADO EN EL USUARIO**

TESIS DOCTORAL

BEATRIZ SAN MIGUEL GONZÁLEZ
Ingeniera de Telecomunicación

2016

**DEPARTAMENTO DE INGENIERÍA
DE SISTEMAS TELEMÁTICOS**

**ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN**

UNIVERSIDAD POLITÉCNICA DE MADRID

**ARQUITECTURA Y MODELO DE REFERENCIA PARA
UN ECOSISTEMA DE INTERCAMBIO DE DATOS
PERSONALES CENTRADO EN EL USUARIO**

TESIS DOCTORAL

AUTORA: BEATRIZ SAN MIGUEL GONZÁLEZ
Ingeniera de Telecomunicación

DIRECTORES: JUAN CARLOS YELMO GARCÍA
Doctor Ingeniero de Telecomunicación

JOSÉ MARÍA DEL ÁLAMO RAMIRO
Doctor por la Universidad Politécnica de Madrid

2016

POLITÉCNICA

Tribunal nombrado por el Magfco. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día.....de.....de 201...

Presidente: _____

Vocal: _____

Vocal: _____

Vocal: _____

Secretario: _____

Suplente: _____

Suplente: _____

Realizado el acto de defensa y lectura de la Tesis el día de de 2016 en la E.T.S.I.T. habiendo obtenido la calificación de _____

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

*A mi madre, por su ejemplo de fuerza y jovialidad
pero, ante todo, de generosidad*

AGRADECIMIENTOS

Una vez me dijeron que la vida del doctorando es un viaje solitario en una montaña rusa donde hay oscuros y claros, dudas y convicciones, complicaciones y oportunidades. No lo pongo en duda, porque he hecho todo el recorrido en el primer vagón. No obstante, esto no hubiera sido posible si no hubieseis estado acompañándome.

En primer lugar, quiero agradecer especialmente a mis directores de tesis su trabajo. A Juan Carlos por la oportunidad y por toda la sabiduría transmitida a lo largo de estos años. A Chema por la ilusión que le ha puesto a esta tesis y porque sin él esto no hubiera sido lo mismo.

Muchas gracias también a todos los compañeros que habéis estado a mi lado. Samuel, compañero de principio a fin, gracias por los consejos y la ayuda. Ya sabes, tenemos al menos dos citas pendientes: una, seguro que en breve, para tu defensa y otra dentro de treinta y cinco años para revisar lo de las pensiones. Rubén, Jorge, Nayeli, Roberto y Cris, gracias porque con vosotros ha sido mucho más llevadero.

A todos mis compañeros del COM. Antonio, Álvaro, Hugo, Paco, Mattia, etc. por los buenos momentos y conversaciones tan entretenidas, realmente me habéis hecho sentir “uno” más. Gracias a July, desde aquí aprovecho para declararme su fan incondicional. Gracias a Juan Carlos Dueñas por dejarme ser testigo de su modelo de gestión impecable.

Sois muchos más los que habéis estado en las distintas etapas, Boni, Sandra, compañeros de Buscamedia y Segur@.

Para todos, confío en que nos vayamos encontrando en lo sucesivo.

No puedo dejar de agradecer a todos aquellos que sin saber muy bien de qué va esto, siempre estáis a mi lado.

Muchas gracias a mi familia. A mis padres, porque esto os llega con la misma intensidad. Gracias por estar siempre ahí y especialmente en este último empujón. A mi hermana Inés, la artista de la familia, porque entendió lo de los modelos casi casi a la vez que yo. Al miembro más reciente, Juan, por aguantar este primer año de convivencia lleno de cambios y siempre con una sonrisa. Gracias por estar a mi lado, tu comprensión y todo tu apoyo.

También a todos mis amigos. A Juanan y Cris, porque siempre estáis dispuestos a todo. Gracias por la compañía y como no, a toda la tropa, Esther, Gorka, Javi, Zoila, Vanesa, Carol, Rodri, Peter, José, etc. por los maravillosos planes que nos montamos. Muchas gracias a la peña el Puntazo y demás allegados aylloneses porque cuando nos juntamos, eso sí que es desconectar. A Carla, mi fiel mujer, porque es maravillosa. A Anita, porque lo suyo da para unas cuantas tesis, impresionante. A Bea C, que superó esta carrera de fondo mucho antes. Gracias Sheila, Miguel, Edu, Tito, Iván, Robert, Raquel, Félix, Emma, etc., tenemos pendientes unas chuletillas en la villa.

Por último, no quiero olvidarme de los que ya no están pero deberían, mi tía Beatriz, Vito y Cipri, veo vuestras caritas de alegría.

RESUMEN

Los servicios telemáticos han transformando la mayoría de nuestras actividades cotidianas y ofrecen oportunidades sin precedentes con características como, por ejemplo, el acceso ubicuo, la disponibilidad permanente, la independencia del dispositivo utilizado, la multimodalidad o la gratuidad, entre otros. No obstante, los beneficios que destacan en cuanto se reflexiona sobre estos servicios, tienen como contrapartida una serie de riesgos y amenazas no tan obvios, ya que éstos se nutren de y tratan con datos personales, lo cual suscita dudas respecto a la privacidad de las personas.

Actualmente, las personas que asumen el rol de usuarios de servicios telemáticos generan constantemente datos digitales en distintos proveedores. Estos datos reflejan parte de su intimidad, de sus características particulares, preferencias, intereses, relaciones sociales, hábitos de consumo, etc. y lo que es más controvertido, toda esta información se encuentra bajo la custodia de distintos proveedores que pueden utilizarla más allá de las necesidades y el control del usuario.

Los datos personales y, en particular, el conocimiento sobre los usuarios que se puede extraer a partir de éstos (modelos de usuario) se han convertido en un nuevo activo económico para los proveedores de servicios. De este modo, estos recursos se pueden utilizar para ofrecer servicios centrados en el usuario basados, por ejemplo, en la recomendación de contenidos, la personalización de productos o la predicción de su comportamiento, lo cual permite a los proveedores conectar con los usuarios, mantenerlos, involucrarlos y en definitiva, fidelizarlos para garantizar el éxito de un modelo de negocio. Sin embargo, dichos recursos también pueden utilizarse para establecer otros modelos de negocio que van más allá de su procesamiento y aplicación individual por parte de un proveedor y que se basan en su comercialización y compartición con otras entidades.

Bajo esta perspectiva, los usuarios sufren una falta de control sobre los datos que les refieren, ya que esto depende de la voluntad y las condiciones impuestas por los proveedores de servicios, lo cual implica que habitualmente deban enfrentarse ante la disyuntiva de ceder sus datos personales o no acceder a los servicios telemáticos ofrecidos. Desde el sector público se trata de tomar medidas que protejan a los usuarios con iniciativas y legislaciones que velen por su privacidad y que aumenten el control sobre sus datos personales, a la vez que debe favorecer el desarrollo económico propiciado por estos proveedores de servicios.

En este contexto, esta tesis doctoral propone una arquitectura y modelo de referencia para un ecosistema de intercambio de datos personales centrado en el usuario que promueve la creación, compartición y utilización de datos personales y modelos de usuario entre distintos proveedores, al mismo tiempo que ofrece a los usuarios las herramientas necesarias para ejercer su control en cuanto a la cesión y uso de sus recursos personales y obtener, en su caso, distintos incentivos o contraprestaciones económicas.

Las contribuciones originales de la tesis son la especificación y diseño de una arquitectura que se apoya en un proceso de modelado distribuido que se ha definido en el marco de esta investigación. Éste se basa en el aprovechamiento de recursos que distintas entidades (fuentes

de datos) ofrecen para generar modelos de usuario enriquecidos que cubren las necesidades específicas de terceras entidades, considerando la participación del usuario y el control sobre sus recursos personales (datos y modelos de usuario). Lo anterior ha requerido identificar y caracterizar las fuentes de datos con potencial de abastecer al ecosistema, determinar distintos patrones para la generación de modelos de usuario a partir de datos personales distribuidos y heterogéneos y establecer una infraestructura para la gestión de identidad y privacidad que permita a los usuarios expresar sus preferencias e intereses respecto al uso y compartición de sus recursos personales. Además, se ha definido un modelo de negocio de referencia que sustenta las investigaciones realizadas y que ha sido particularizado en dos ámbitos de aplicación principales, en concreto, el sector de publicidad en redes sociales y el sector financiero para la implantación de nuevos servicios.

Finalmente, cabe destacar que las contribuciones de esta tesis han sido validadas en el contexto de distintos proyectos de investigación industrial aplicada y también en el marco de proyectos fin de carrera que la autora ha tutelado o en los que ha colaborado. Los resultados obtenidos han originado distintos méritos de investigación como dos patentes en explotación, la publicación de un artículo en una revista con índice de impacto y diversos artículos en congresos internacionales de relevancia. Algunos de éstos han sido galardonados con premios de distintas instituciones, así como en las conferencias donde han sido presentados.

ABSTRACT

Information society services have changed most of our daily activities, offering unprecedented opportunities with certain characteristics, such as: ubiquitous access, permanent availability, device independence, multimodality and free-of-charge services, among others. However, all the positive aspects that emerge when thinking about these services have as counterpart not-so-obvious threats and risks, because they feed from and use personal data, thus creating concerns about peoples' privacy.

Nowadays, people that play the role of user of services are constantly generating digital data in different service providers. These data reflect part of their intimacy, particular characteristics, preferences, interests, relationships, consumer behavior, etc. Controversy arises because this personal information is stored and kept by the mentioned providers that can use it beyond the user needs and control.

Personal data and, in particular, the knowledge about the user that can be obtained from them (user models) have turned into a new economic asset for the service providers. In this way, these data and models can be used to offer user centric services based, for example, in content recommendation, tailored-products or user behavior, all of which allows connecting with the users, keeping them more engaged and involved with the provider, finally reaching customer loyalty in order to guarantee the success of a business model. However, these resources can be used to establish a different kind of business model; one that does not only processes and individually applies personal data, but also shares and trades these data with other entities.

From that perspective, the users lack control over their referred data, because it depends from the conditions imposed by the service providers. The consequence is that the users often face the following dilemma: either giving up their personal data or not using the offered services. The Public Sector takes actions in order to protect the users approving, for example, laws and legal initiatives that reinforce privacy and increase control over personal data, while at the same time the authorities are also key players in the economy development that derives from the information society services.

In this context, this PhD Dissertation proposes an architecture and reference model to achieve a user-centric personal data ecosystem that promotes the creation, sharing and use of personal data and user models among different providers, while offering users the tools to control who can access which data and why and if applicable, to obtain different incentives.

The original contributions obtained are the specification and design of an architecture that supports a distributed user modelling process defined by this research. This process is based on leveraging scattered resources of heterogeneous entities (data sources) to generate on-demand enriched user models that fulfill individual business needs of third entities, considering the involvement of users and the control over their personal resources (data and user models). This has required identifying and characterizing data sources with potential for supplying resources, defining different generation patterns to produce user models from scattered and heterogeneous data, and establishing identity and privacy management

infrastructures that allow users to set their privacy preferences regarding the use and sharing of their resources. Moreover, it has also been proposed a reference business model that supports the aforementioned architecture and this has been studied for two application fields: social networks advertising and new financial services.

Finally, it has to be emphasized that the contributions obtained in this dissertation have been validated in the context of several national research projects and master thesis that the author has directed or has collaborated with. Furthermore, these contributions have produced different scientific results such as two patents and different publications in relevant international conferences and one magazine. Some of them have been awarded with different prizes.

TABLA DE CONTENIDOS

RESUMEN.....	I
ABSTRACT	III
TABLA DE CONTENIDOS.....	V
LISTA DE FIGURAS.....	IX
LISTA DE TABLAS.....	XI
GLOSARIO	XIII
CAPÍTULO 1. INTRODUCCIÓN	1
1.1 ESTRUCTURA DEL DOCUMENTO DE LA TESIS DOCTORAL	4
1.2 METODOLOGÍA DE INVESTIGACIÓN	6
CAPÍTULO 2. OBJETIVOS	7
2.1 OBJETIVOS ESPECÍFICOS	7
CAPÍTULO 3. ESTADO DEL ARTE	9
3.1 DEFINICIONES Y TERMINOLOGÍA.....	9
3.1.1 Datos personales.....	10
3.1.2 Perfil y modelo de usuario	13
3.1.3 Identidad digital	16
3.1.4 Mapa conceptual de la tesis doctoral	17
3.2 PROCESO DE MODELADO DE USUARIO	18
3.2.1 Recolección de datos personales	19
3.2.2 Generación de modelos de usuario	21
3.2.3 Aplicación del modelo de usuario y realimentación	24
3.2.4 Otros trabajos relacionados.....	25
3.3 MECANISMOS DE RECUPERACIÓN DE DATOS PERSONALES ALMACENADOS EN FUENTES DE DATOS EXTERNAS	26
3.3.1 Programas araña	27
3.3.2 Ficheros estáticos de datos.....	28
3.3.3 Mecanismos de petición	28
3.3.4 Mecanismos de suscripción	31
3.3.5 Resumen	34
3.4 GESTIÓN DE IDENTIDAD	34
3.4.1 Modelos de gestión de identidad	35
3.4.2 Tecnologías para la delegación de autorización	43
3.4.3 Resumen	45
3.5 PRIVACIDAD	45
3.5.1 Privacidad desde el diseño.....	48

3.5.2	Ingeniería de la privacidad	49
3.5.3	Resumen	52
3.6	MODELOS DE NEGOCIO EN EL ECOSISTEMA DE DATOS PERSONALES ACTUAL	52
3.6.1	Concepto y elementos del modelo de negocio.....	53
3.6.2	Modelos de negocio en el contexto de servicios telemáticos	53
3.6.3	Modelos de negocio basados en datos personales	58
3.6.4	Resumen	62
CAPÍTULO 4. NUEVOS MODELOS DE NEGOCIO PARA EL ECOSISTEMA DE DATOS PERSONALES.....		63
4.1	POSICIONAMIENTO	64
4.2	MODELO DE NEGOCIO DE REFERENCIA PARA UN ECOSISTEMA DE INTERCAMBIO DE DATOS PERSONALES CENTRADO EN EL USUARIO	67
4.2.1	Actores, roles y relaciones	67
4.2.2	Propuestas de valor	70
4.2.3	Segmentación del mercado	71
4.2.4	Cadena de valor	72
4.2.5	Red de valor	73
4.2.6	Mecanismos de generación de ingresos y otros incentivos.....	75
4.3	PARTICULARIZACIÓN DEL MODELO DE NEGOCIO EN DISTINTOS ÁMBITOS DE APLICACIÓN ...	76
4.3.1	Modelo de negocio de intermediación de publicidad en redes sociales	77
4.3.2	Modelo de negocio para la promoción de nuevos servicios financieros	83
4.4	RESUMEN Y CONTRIBUCIONES DEL CAPÍTULO	88
CAPÍTULO 5. ESPECIFICACIÓN Y DISEÑO DE UN MARCO DE DATOS PERSONALES DE REFERENCIA		91
5.1	ESPECIFICACIÓN DEL MARCO DE DATOS PERSONALES.....	92
5.1.1	Descripción funcional de alto nivel mediante casos de uso	93
5.1.2	Especificación complementaria: requisitos no funcionales.....	98
5.2	ARQUITECTURA DE REFERENCIA	101
5.2.1	Gestor de usuarios	102
5.2.2	Recolector	103
5.2.3	Registro	103
5.2.4	Generador	104
5.2.5	Almacén de recursos personales	105
5.2.6	Recuperador	105
5.3	RESUMEN Y CONTRIBUCIONES DEL CAPÍTULO	106
5.3.1	Enfoque de validación global	107
CAPÍTULO 6. RECUPERACIÓN DE DATOS EN EL MARCO DE DATOS PERSONALES		109
6.1	IDENTIFICACIÓN DE FUENTES DE DATOS EXTERNAS.....	110
6.2	CARACTERIZACIÓN DE LAS FUENTES DE DATOS EXTERNAS.....	112
6.2.1	El usuario como fuente de datos	114

6.2.2	Proveedores de servicios como fuentes de datos personales	115
6.2.3	Otras fuentes de datos y recursos disponibles	121
6.2.4	Resumen de las características de las fuentes de datos externas	124
6.3	RECUPERACIÓN DE RECURSOS EN EL MARCO DE DATOS PERSONALES	126
6.4	VALIDACIÓN DE COMPONENTES ESPECÍFICOS PARA LA RECUPERACIÓN DE DATOS.....	130
6.4.1	Recuperación de datos personales en servicios de mensajería y redes sociales: el servicio de pago Posdata.....	130
6.4.2	Producción y recuperación de eventos personales de negocio en una arquitectura empresarial del ámbito bancario	138
6.5	RESUMEN Y CONTRIBUCIONES DEL CAPÍTULO	146
CAPÍTULO 7. GENERACIÓN Y APLICACIÓN DE MODELOS DE USUARIO		149
7.1	CARACTERÍSTICAS DE LOS MODELOS DE USUARIO DEL MARCO DE DATOS PERSONALES	150
7.2	PATRONES PARA LA GENERACIÓN DE MODELOS DE USUARIO	152
7.2.1	Patrón centralizado.....	153
7.2.2	Patrón distribuido	155
7.2.3	Patrón desacoplado	157
7.2.4	Patrones híbridos	158
7.2.5	Patrones de generación y características de los modelos de usuario asociados.....	159
7.3	VALIDACIONES DE LOS PATRONES DE GENERACIÓN PARA LA CREACIÓN DE NUEVOS SERVICIOS FINANCIEROS	161
7.3.1	Patrón centralizado para generación de modelos de usuario socio-económicos	161
7.3.2	Patrón desacoplado y distribuido para la generación de modelos de usuario enriquecidos de corto plazo	170
7.4	RESUMEN Y CONTRIBUCIONES DEL CAPÍTULO	175
CAPÍTULO 8. GESTIÓN DE IDENTIDAD Y PRIVACIDAD CENTRADA EN EL USUARIO		177
8.1	INFRAESTRUCTURA PARA LA GESTIÓN DE LA IDENTIDAD.....	179
8.1.1	Modelos de gestión de identidad	180
8.1.2	Arquitectura para la gestión de identidad	182
8.2	INFRAESTRUCTURA DE GESTIÓN DE LA PRIVACIDAD.....	184
8.2.1	Gestión de privacidad en el marco de datos personales	185
8.2.2	Gestión de privacidad desde el punto de vista del usuario	189
8.3	VALIDACIONES DE LAS INFRAESTRUCTURAS DE GESTIÓN DE IDENTIDAD Y PRIVACIDAD... 	191
8.3.1	Gestión de identidades digitales en servicios de redes sociales.....	191
8.3.2	Controlador de privacidad para un dominio federado de gestión de identidad	193
8.4	RESUMEN Y CONTRIBUCIONES DEL CAPÍTULO	197
CAPÍTULO 9. VALIDACIÓN Y RESULTADOS.....		199
9.1	VALIDACIONES EN PROYECTOS DE INVESTIGACIÓN Y OTROS TRABAJOS ACADÉMICOS	200
9.1.1	Validación en el proyecto Segur@	200
9.1.2	Validación en el proyecto Buscamedia	201

9.1.3	Validación en el Center for Open Middleware	203
9.1.4	Validación con soporte de trabajos académicos.....	205
9.2	RESULTADOS OBTENIDOS.....	206
9.2.1	Protección intelectual e industrial a través de patentes	206
9.2.2	Diseminación de resultados.....	207
9.2.3	Premios otorgados.....	209
9.3	RESUMEN DEL CAPÍTULO.....	210
CAPÍTULO 10. CONCLUSIONES		211
10.1	CONSECUCCIÓN DE LOS OBJETIVOS ORIGINALES DE LA TESIS	212
10.2	CONTRIBUCIONES DE LA TESIS DOCTORAL	215
10.3	TRABAJO FUTURO.....	217
REFERENCIAS.....		219

LISTA DE FIGURAS

Figura 1-1. Datos creados por los usuarios en los principales proveedores de servicios de Internet (DOMO, 2015).	2
Figura 3-1. Taxonomía de datos personales (Halim, 2011).	13
Figura 3-2. Tipos de modelos de usuario (Asif, 2014).	16
Figura 3-3. Mapa conceptual de la tesis doctoral (San Miguel, et al., 2014c).	18
Figura 3-4. Etapas principales del modelado de usuario (Barla, 2010) y (San Miguel, et al., 2015c).	19
Figura 3-5. Ejemplo de aviso sobre la utilización de cookies.	20
Figura 3-6. Visión global del ecosistema de economía API (San Miguel, et al., 2015b).	29
Figura 3-7. Niveles de una arquitectura dirigida a eventos genérica (San Miguel, et al., 2014b).	32
Figura 3-8. Modelo de gestión de identidad aislado (Jøsang y Pope, 2005).	36
Figura 3-9. Modelo de gestión de identidad federado (Jøsang y Pope, 2005).	37
Figura 3-10. Gestión de identidades federada de las especificaciones del proyecto Liberty (Yelmo, et al., 2009).	38
Figura 3-11. Modelo de gestión de identidad centralizado (Jøsang y Pope, 2005).	39
Figura 3-12. Procedimiento de autenticación en OpenID (Martin y Gil, 2012).	40
Figura 3-13. Ejemplo de registro a través del <i>social login</i> .	41
Figura 3-14. Proveedores de servicios elegidos por los usuarios para la autenticación en dispositivos móviles (parte izquierda) y la realización de compras electrónicas (parte derecha) en el segundo cuatrimestre del año 2015 (Gigya, 2015).	41
Figura 3-15. Ejemplo de un sistema de gestión de identidad centrado en el usuario (Jøsang y Pope, 2005).	42
Figura 3-16. Gestión de identidad en el proveedor de servicios Facebook.	43
Figura 3-17. Visión global del protocolo OAuth (San Miguel, et al., 2015b).	44
Figura 3-18. Visión global del enfoque de privacidad desde el diseño (Privacy by Design, 2011).	48
Figura 3-19. Arquitectura y secuencia para la aplicación de políticas de privacidad (Kim, 2009).	52
Figura 3-20. Modelo de negocio de los brókeres de datos personales.	59
Figura 3-21. Modelo de negocio de publicidad.	60
Figura 3-22. Modelo de negocio de plataformas abiertas de datos.	61
Figura 4-1. Visión global del modelo de negocio del marco de datos personales.	66
Figura 4-2. Cadena de valor del modelo de negocio del marco de datos personales.	73
Figura 4-3. Red de valor para el marco de datos personales.	74
Figura 4-4. Lanzamiento de distintas redes sociales (1997-2010) (Fundación Telefónica, 2011).	78
Figura 4-5. Visión global del modelo de negocio del OSAM.	80
Figura 4-6. Cadena de valor del modelo de negocio del OSAM.	82
Figura 4-7. Visión global del modelo de negocio del bróker financiero.	86
Figura 5-1. Visión global para la especificación de un marco de datos personales de referencia.	92
Figura 5-2. Diagrama de casos de uso del marco de datos personales de referencia.	93
Figura 5-3. Diagrama de casos de uso extendido del marco de datos personales de referencia.	94
Figura 5-4. Arquitectura de referencia del marco de datos personales (San Miguel, et al., 2015a).	102
Figura 6-1. Proceso de modelado de usuario distribuido en el marco de datos personales.	110
Figura 6-2. Visión global de la recuperación de recursos en el marco de datos personales.	128
Figura 6-3. Interfaz de usuario de activación del servicio Posdata (San Miguel, et. al, 2015b).	134
Figura 6-4. Publicación de Facebook con la realización de un pago en el servicio Posdata (San Miguel, et. al, 2014c).	135
Figura 6-5. Arquitectura del servicio Posdata.	136
Figura 6-6. Proceso de recuperación de datos del servicio Posdata.	137

Figura 6-7. Entorno de ejecución de BankSphere (San Miguel, et al., 2014b).	140
Figura 6-8. Definición de los eventos de una arquitectura empresarial (San Miguel, et al., 2015d).	142
Figura 6-9. Arquitectura de referencia para la producción y recuperación de eventos personales de negocio en BankSphere.	143
Figura 7-1. Modelos de usuario considerados en el marco de datos personales.	152
Figura 7-2. Patrón centralizado para la generación de modelos de usuario.	154
Figura 7-3. Patrón distribuido para la generación de modelos de usuario.	156
Figura 7-4. Patrón desacoplado para la generación de modelos de usuario.	157
Figura 7-5. Escenario de validación para el patrón centralizado de generación de modelos de usuario (San Miguel, et al., 2015c).	162
Figura 7-6. Resumen ontología Personal Socio-Economic Network (San Miguel, et al., 2015a).	164
Figura 7-7. Elementos para el razonamiento basado en reglas (Flores Vitelli, 2011).	166
Figura 7-8. Captura de pantalla del servicio FriendLoans (San Miguel, et al., 2015c).	168
Figura 7-9. Escenario de validación para los patrones desacoplado y distribuido de generación de modelos los de usuario.	171
Figura 7-10. Herramienta de visualización para el consumo de eventos personales de negocio e información relacionada.	174
Figura 8-1. Diagrama de flujos de los datos personales en el contexto de la tesis.	178
Figura 8-2. Infraestructura de gestión de identidad en el marco de datos personales.	181
Figura 8-3. Políticas de privacidad en el contexto del marco de datos personales.	185
Figura 8-4. Gestión e implantación de políticas en el marco de datos personales.	187
Figura 8-5. Diagrama de secuencia para la presentación de recursos vinculados.	189
Figura 8-6. Diagrama de secuencia para la modificación/borrado recurso personal.	190
Figura 8-7. Diagrama de secuencia para la consulta de historial de uso.	191
Figura 8-8. Controlador de privacidad en un dominio federado.	193
Figura 8-9. Diagrama de secuencia para la presentación de recursos en dominio federado.	194
Figura 8-10. Diagrama de secuencia para la visualización de valores asociados a un recurso y su modificación.	195
Figura 8-11. Interfaz del controlador de privacidad con el historial de uso de la tarjeta de crédito de un usuario.	196
Figura 8-12. Ejemplo interfaz de usuario para el establecimiento de políticas de privacidad.	197

LISTA DE TABLAS

Tabla 4-1. Relaciones entre los actores y roles del modelo de negocio.....	69
Tabla 4-2. Resumen de los ingresos del modelo de negocio del marco de datos personales.	76
Tabla 6-1. Fuentes de datos externas identificadas y utilidad en el proceso de modelado de usuario...112	
Tabla 6-2. Resumen de las características de los proveedores de servicios como fuente de datos.	125
Tabla 7-1. Patrones de generación y características de los modelos de usuario asociados.	160

GLOSARIO

AJAX	Asynchronous JavaScript And XML
API	Application Programming Interfaces
BKS	BankSphere
CDTI	Centro para el Desarrollo Tecnológico Industrial
CENIT	Consorcios Estratégicos Nacionales en Investigación Técnica
CLIPS	C Language Integrated Production System
CNMC	Comisión Nacional de los Mercados y la Competencia
COM	Center for Open Middleware
CRM	Customer Relationship Management
CSIC	Consejo Superior de Investigaciones Científicas
CSV	Comma-Separated Values
DS	Discovery Service
EDA	Event-Driven Architecture
ENISA	European Union Agency for Network and Information Security
FIBO	Financial Industry Business Ontology
FOAF	Friend of a Friend
FpML	Financial products Markup Language
FTC	Federal Trade Commission
HCI	Human–computer interaction
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
IdP	Identity Provider
IEEE	Instituto de Ingeniería Eléctrica y Electrónica
IETF	Internet Engineering Task Force
INE	Instituto Nacional de Estadística
ISO	International Organization for Standardization

ITU	International Telecommunication Union
JSON	JavaScript Object Notation
MOM	Message Oriented Middleware
NEM	Networked & Electronic Media
NSA	Agencia Nacional de Seguridad de los Estados Unidos
OASIS	Organization for the Advancement of Structured Information Standards
OECD	Organisation for Economic Cooperation and Development OCDE
OMG	Object Management Group
ONTSI	Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información
OSAM	Online Social Advertisement Mediator
OWL	Web Ontology Language
P3P	Platform for Privacy Preferences Project
PbD	Privacy by Design
PeDF	Personal Data Framework
PEST	Análisis de factores Político, Económico, Social y Tecnológico
PET	Privacy Enhancing Technologies
PII	Personally Identifiable Information
PSEN	Personal Socio-Economic Network
RDF	Resource Description Framework
REST	Representational State Transfer
SOA	Service Oriented Architecture
SP	Service Provider
SSO	Single Sign-On
SUMO	Suggested Upper Merged Ontology
TIC	Tecnologías de la Información y Comunicación
TUI	Tarjeta Universitaria Inteligente
URL	Uniform Resource Locator
W3C	World Wide Web

WEF World Economic Forum

WIPO World Intellectual Property Organization

XACML eXtensible Access Control Markup Language

XBRL eXtensible Business Reporting Language

XML eXtensible Markup Language

XSD XML Schema Definition

Capítulo 1.

Introducción

La sociedad del siglo XXI ha sido denominada como la *sociedad red* (Castells, 2006), dado que ésta se ha construido en torno a las Tecnologías de la Información y Comunicación (TIC) y en particular, a Internet y los servicios telemáticos. Dichas tecnologías definen la estructura social actual y su importancia es tal que incluso los individuos que no las utilizan se ven afectados por las interacciones del resto de entidades participantes.

Los servicios telemáticos han irrumpido en nuestras vidas de una forma sorprendente, transformando la mayoría de nuestras actividades cotidianas relacionadas con aspectos tan relevantes como el empleo, la educación, las relaciones sociales, el consumo o el entretenimiento, por citar sólo algunos ejemplos. Actualmente, estos servicios se han convertido en una herramienta fundamental que ofrece a los usuarios oportunidades sin precedentes, puesto que permiten realizar multitud de tareas dispares desde cualquier lugar y en cualquier momento. No obstante, los beneficios y oportunidades que destacan en cuanto se reflexiona sobre dichos servicios, tienen como contrapartida una serie de riesgos y amenazas no tan obvios, ya que dichos servicios se nutren de y tratan con datos personales de los usuarios, lo cual suscita dudas y preocupaciones respecto a la privacidad de las personas.

El simple gesto de registrarse en un servicio, y aún más su uso continuado, provoca que los usuarios creen y generen datos personales, trasladando parte de su intimidad y vida privada, en diferentes proveedores de servicios y, lo que es más controvertido, a merced de éstos. Los usuarios se han acostumbrado al acceso ubicuo, la disponibilidad permanente, la independencia del dispositivo utilizado, la multimodalidad y habitualmente, al uso gratuito de los servicios telemáticos y bien por desconocimiento o por la compensación de utilizar estos servicios, acceden a la cesión de sus datos personales a proveedores de servicios.

Es importante destacar en este punto que, tal y como señalan distintos informes del Eurobarómetro (Comisión Europea, 2011) y (Comisión Europea, 2015) u otras organizaciones independientes (Madden y Rainie, 2015), los usuarios de Internet muestran su inquietud y preocupación por la utilización, la falta de control y compartición de sus datos personales. Sin embargo, en este sentido, lo que se ha venido en llamar la *paradoja de la privacidad* (Barnes,

2006) refleja que los usuarios, aun estando preocupados por el uso de sus datos personales y declarar como inadmisibles ciertas conductas, en situaciones reales ceden ante los beneficios que obtienen y siguen proporcionando y compartiendo sus datos personales en proveedores de servicios.

Ante la situación planteada, la realidad es que el volumen de datos personales creados por los usuarios y obtenidos por los proveedores de servicios va en aumento y la previsión futura es de crecimiento (Gantz y Reinsel, 2012). A modo de ejemplo, y tomando como referencia el análisis realizado por la empresa DOMO¹ cuyo resumen puede apreciarse en la Figura 1-1, cada minuto del día los usuarios indican que les gustan más de 4 millones de publicaciones en Facebook², se envían casi 350 mil mensajes en Twitter³, se suben 300 horas de nuevos videos en Youtube⁴, etc.

Figura 1-1. Datos creados por los usuarios en los principales proveedores de servicios de Internet (DOMO, 2015).

El aprovechamiento de todos estos datos y en particular, su monetización, es una realidad del nuevo contexto. Tal y como expresó la Comisaria Europea de Consumo (Kuneva, 2009), los datos personales son el nuevo petróleo de Internet y la nueva moneda de cambio del contexto digital. A título ilustrativo, grandes proveedores de servicios, como Google⁵ o Facebook, basan sus exitosos modelos de negocio en la explotación de datos personales, comercializando con el conocimiento que obtienen de sus usuarios, sin que éstos sean completamente conscientes, todo a cambio de ofrecerles servicios muy atractivos aparentemente gratuitos.

Dadas las consideraciones anteriores y conforme a lo expuesto por el Foro Económico Mundial (*World Economic Forum* – WEF, 2011), los datos personales de los usuarios se han convertido en el nuevo activo económico del siglo XXI y han propiciado la **configuración de un nuevo**

¹ DOMO, <https://www.domo.com/>

² Facebook, <http://facebook.com/>

³ Twitter, <http://twitter.com/>

⁴ YouTube, <https://www.youtube.com/>

⁵ Google, <https://www.google.com/>

ecosistema que presenta desigualdades y en el que los usuarios, el sector privado y el sector público se enfrentan ante los aspectos que se detallan a continuación.

Como ya se ha destacado anteriormente, los **usuarios** se encuentran ante la disyuntiva de ceder sus datos personales o no acceder a servicios telemáticos, en definitiva, sin un control real sobre los datos que les refieren, puesto que esto depende de la voluntad y las condiciones impuestas por los proveedores de servicios.

Desde la perspectiva del **sector privado**, la provisión de servicios centrados en el usuario, basados en ofrecer una experiencia de uso adaptada a las preferencias y necesidades personales de los usuarios, se ha convertido en un área primordial que requiere la obtención de datos personales y conocimiento muy preciso de los usuarios. Es decir, los datos personales se han convertido en el eje principal para la provisión de servicios y además, su posición frente a ellos determinan muchos de los modelos de negocio actuales. En este sentido, han aparecido modelos de negocio que van más allá de su procesamiento y aplicación individual por parte de un proveedor y que se basan en su comercialización y compartición con otras entidades, lo cual puede ser considerado como una invasión a la privacidad de las personas.

Desde el **sector público**, distintos gobiernos, instituciones y organizaciones tienen que hacer frente a las situaciones descritas. De esta forma, bajo el sentimiento generalizado de que éticamente no todo es posible, tratan de proteger a los usuarios (ciudadanos) con legislaciones que velen por un uso correcto y adecuado de sus datos personales y en definitiva, que protejan su privacidad, a la vez que ejercen como responsables de la promoción de iniciativas que permitan la generación de nuevo valor y el desarrollo económico.

No obstante, existen diferencias territoriales con respecto al enfoque regulatorio de los datos personales. Por una parte, Europa declara la protección de los datos personales y el respeto de la vida privada como derechos fundamentales y con el objeto de garantizarlos cuenta con distintos instrumentos como, por ejemplo, la Directiva 95/46/CE relativa a la protección de datos o la Directiva 2002/58/CE, modificada en 2009, sobre la privacidad y las comunicaciones electrónicas, entre otros. Además, Europa está inmersa en una nueva propuesta legislativa que trata de salvaguardar los datos personales en todo el territorio de la Unión Europea, aumentar el control de los datos por parte de los usuarios y reducir los costes para las empresas (Davoli, 2015).

Por otra parte, otros países, como es el caso de los Estados Unidos, se han posicionado con enfoque basado en la autorregulación donde el sector privado actúa conforme a sus necesidades aunque existen excepciones como, por ejemplo, aquellas relacionadas con el uso de datos bancarios, médicos o de menores, y cada vez más se están posicionando con iniciativas que promueven la cultura de la privacidad y la elaboración de códigos de conducta que apoyan a la industria a elaborar sus esquemas de autorregulación (*Federal Trade Commission – FTC*, 2015).

Ante el contexto planteado, esta tesis doctoral propone una arquitectura y modelo de referencia para un ecosistema de intercambio de datos personales centrado en el usuario que promueve la creación, compartición y utilización de datos personales y conocimiento de los usuarios a través de modelos de usuario entre distintos proveedores, al mismo tiempo que

ofrece a los usuarios las herramientas necesarias para establecer su control en cuanto a la cesión y uso de sus recursos personales y la posibilidad de recibir distintas contraprestaciones económicas u otros incentivos.

Para lo anterior, se abordará el estudio y la combinación de tres áreas de investigación que se originan en torno a los datos personales de los usuarios: modelado de usuario, gestión de identidad y gestión de privacidad, y que habitualmente se proyectan como áreas independientes. Además, se estudiarán los distintos modelos de negocio que actualmente se basan en la compartición de datos personales entre proveedores de servicios para establecer un modelo de referencia. Cada una de las contribuciones originales de la tesis se validará en distintos ámbitos donde su aplicación es especialmente relevante. En concreto, se presentarán escenarios de validación relacionados con servicios de redes sociales y nuevos servicios financieros que han sido definidos en proyectos de investigación industrial aplicada.

1.1 Estructura del documento de la tesis doctoral

Una vez realizada la introducción y presentado el contexto en el que se enmarca la tesis doctoral, el documento se organiza como sigue:

- **Capítulo 2 Objetivos.** Esboza el objetivo general de la tesis doctoral y, sobre la base de éste, expone los objetivos específicos de la misma.
- **Capítulo 3 Estado del arte.** Incluye un análisis de las áreas de investigación, soluciones y tecnologías que permiten comprender y apoyar las contribuciones presentadas en esta tesis doctoral. Para ello, en primer lugar, se presentan la terminología, las definiciones más relevantes de esta investigación y un mapa conceptual de las áreas de investigación relacionadas con la tesis: modelado de usuario, gestión de identidad y privacidad. Seguidamente, se detalla el estado del arte relacionado con dichas áreas y para finalizar el capítulo, se incluye un estudio de los modelos de negocio que se establecen en el contexto de servicios telemáticos y en particular, en aquellos que se basan en el uso e intercambio de datos personales entre proveedores de servicios.
- **Capítulo 4 Nuevos modelos de negocio para el ecosistema de datos personales.** En este capítulo se expone el posicionamiento establecido en esta investigación frente a los modelos de negocio que se han originado en torno a la utilización y compartición de datos personales. Sobre la base de esto, se establece un modelo de negocio de referencia que se describe resaltando sus elementos principales (actores participantes, roles que asumen éstos, propuestas, cadena y red de valor, etc.) y dos particularizaciones de dicho modelo (una para el sector de publicidad en redes sociales y otra para el sector financiero) que se han llevado a cabo en esta tesis doctoral.
- **Capítulo 5 Especificación y diseño de un marco de datos personales de referencia.** Presenta la especificación y diseño de una arquitectura de referencia que posibilita el modelo de negocio descrito en el punto anterior. De esta forma, en este capítulo se formaliza la especificación de un marco de datos personales, considerando los servicios que éste debe incorporar y los principios fundamentales en los que el sistema

debe estar fundamentado y que determinan su correcto funcionamiento. Además, tal y como se señalará en la última sección del capítulo, se presenta el enfoque de validación global de dicha arquitectura y que se describe en los capítulos posteriores 6, 7 y 8.

- **Capítulo 6 Recuperación de datos en el marco de datos personales.** Aborda uno de los pilares fundamentales de la arquitectura de referencia establecida: la recuperación de recursos, que permitirá posteriormente llevar a cabo la generación y aplicación de modelos de usuario. En este capítulo se identifican y caracterizan distintas fuentes de datos personales con potencial de abastecer a la arquitectura establecida de los recursos necesarios. Además, se propone un modelo global para la recuperación de los recursos identificados y por último, se presentan dos validaciones de las contribuciones originales incluidas. Dichas validaciones permiten la recuperación de datos personales en redes sociales y servicios de mensajería y la obtención de un tipo de datos que se ha identificado en el contexto de esta investigación y que se origina dentro de una arquitectura empresarial, en particular, del ámbito bancario.
- **Capítulo 7 Generación y aplicación de modelos de usuario.** En este capítulo se establecen distintos patrones que permiten la generación de modelos de usuario a partir de recursos heterogéneos y distribuidos obtenidos conforme a lo establecido en el capítulo anterior. Además, se incluye la validación de estos patrones en distintos escenarios financieros establecidos para la promoción de nuevos servicios y la mejora de servicios bancarios existentes.
- **Capítulo 8 Gestión de identidad y privacidad centrada en el usuario.** Este capítulo aborda el último pilar considerado dentro de la arquitectura de referencia establecida en el capítulo 5: las infraestructuras para la gestión de identidad y la gestión de privacidad. Estas infraestructuras sustentan y permiten la realización de la recuperación de datos descrita en el capítulo 6 y la generación y aplicación de modelos de usuarios descritos en el capítulo 7. Para ello se sigue un enfoque centrado en el usuario donde éste es el centro de todos los procesos establecidos en los capítulos anteriores.
- **Capítulo 9 Validación y resultados.** Resume los trabajos de validación realizados en esta tesis doctoral destacando los proyectos de investigación en los que se ha participado, así como los méritos y resultados obtenidos.
- **Capítulo 10 Conclusiones.** Analiza los resultados obtenidos conforme a los objetivos establecidos para la tesis doctoral y señala distintas líneas de trabajo futuro.

En definitiva, se incluyen tres capítulos que introducen y sintetizan el contexto, objetivos y el estado del arte relacionado con la tesis doctoral (capítulos 1, 2 y 3), cinco capítulos con contribuciones originales de la misma (capítulos 4, 5, 6, 7 y 8) y dos capítulos finales que recapitulan los trabajos realizados, resultados obtenidos y las conclusiones extraídas en esta investigación (capítulos 9 y 10).

En último lugar, se encuentran las referencias bibliográficas utilizadas.

1.2 Metodología de investigación

La metodología de investigación utilizada en esta tesis doctoral se apoya en un método inductivo que consiste en la observación y abstracción de resultados obtenidos en distintos ámbitos de aplicación que posteriormente han sido generalizados para obtener un marco y modelo general que ha sido trasladado a otros escenarios de validación para comprobar su validez.

Dicha metodología se apoya en la realización de seis etapas:

1. Identificación, estudio y análisis del estado del arte asociado.
2. Definición y establecimiento de los distintos ámbitos de aplicación de la tesis.
3. Diseño e implementación de técnicas de modelado de usuario, gestión de identidad y privacidad en los ámbitos establecidos.
4. Observación de los resultados obtenidos en los ámbitos de aplicación establecidos en el punto 2.
5. Abstracción de los resultados comunes obtenidos en los distintos ámbitos de aplicación para definir un marco y modelo general.
6. Validación del marco y modelo general establecido en otros escenarios que permitan demostrar su validez.

Capítulo 2.

Objetivos

Conforme a lo presentado en la introducción de este documento (Capítulo 1), el ecosistema de datos personales actual presenta una serie de desigualdades y situaciones controvertidas que requieren tomar medidas específicas en la dimensión social, económica, legal y tecnológica para su desenlace. Para contribuir a tal efecto, esta tesis doctoral aborda el establecimiento de un ecosistema de intercambio de datos personales centrado en el usuario que promueva la reconciliación entre las posturas frecuentemente enfrentadas de proveedores de servicios y usuarios respecto al uso y tratamiento de datos personales. Este ecosistema debe venir definido por una arquitectura y modelo de negocio de referencia que se sitúen entre distintos proveedores de servicios para permitir la creación, compartición y utilización de datos personales de los usuarios y conocimiento sobre éstos con objeto de impulsar sus modelos de negocio particulares, a la vez que involucre y ofrezca a los usuarios los instrumentos necesarios para controlar el flujo de su información personal y obtener, en su caso, distintas contraprestaciones económicas u otros incentivos.

De esta forma, el objetivo general de la tesis doctoral es la especificación, diseño y validación de una arquitectura para el establecimiento de un marco de datos personales que integra la información ofrecida por proveedores de servicios heterogéneos para facilitar la obtención y descubrimiento de nuevo conocimiento del usuario, que difícilmente puede darse de forma unilateral, a otros proveedores de servicios más desfavorecidos. Además, esta arquitectura debe incluir los mecanismos necesarios para permitir a los usuarios comprender el flujo de su información personal y así, establecer sus criterios y preferencias de privacidad en cuanto a la cesión y uso de la misma. Conjuntamente, lo anterior debe estar sustentando sobre un modelo de negocio coherente con la realidad actual que pueda ser trasladado a distintos ámbitos de aplicación.

2.1 Objetivos específicos

Sobre la base del objetivo principal enunciado anteriormente, se derivan los siguientes objetivos específicos para esta tesis doctoral:

- Identificar y analizar los principales modelos de negocio aplicables dentro del ecosistema de datos personales actual con el objeto de proponer un nuevo modelo en torno a la utilización e intercambio de datos personales.
- Investigar y analizar las áreas de investigación y tecnologías asociadas relacionadas con la obtención de conocimiento y la gestión de datos personales de los usuarios.
- Especificar y diseñar una arquitectura de referencia que posibilite la realización del modelo de negocio identificado. Esta arquitectura debe apoyarse en el abastecimiento de recursos externos, la generación de conocimiento de los usuarios, la provisión de éste a terceras entidades y la incorporación de herramientas centradas en el usuario para la gestión de su privacidad.
- Identificar y caracterizar las fuentes de datos personales con potencial de abastecer de los recursos necesarios a la arquitectura de referencia establecida en el punto anterior.
- Determinar distintos modelos para la generación de conocimiento de los usuarios a partir de recursos distribuidos y heterogéneos. En particular, se deberán explorar soluciones basadas en tecnologías semánticas.
- Establecer una infraestructura de gestión de identidad y privacidad centrada en el usuario que permita habilitar la gestión del flujo de sus datos personales para establecer el control en el uso y la cesión de éstos.
- Validar los resultados obtenidos a través de distintos prototipos y contribuciones en el ámbito de proyectos de investigación industrial aplicada.
- Diseminar los resultados obtenidos en revistas y congresos internacionales relevantes de temática relacionada, así como adoptar las medidas necesarias para proteger su propiedad intelectual e industrial a través de patentes.

Capítulo 3.

Estado del arte

La presente tesis doctoral se establece en torno a tres áreas de investigación principales que se originan respecto al uso e intercambio de datos personales: modelado de usuario, gestión de la identidad y privacidad. Además, aborda una nueva perspectiva de negocio respecto a la gestión y compartición de los datos personales de los usuarios.

En este capítulo se pone en contexto al lector respecto a los aspectos señalados, destacando las soluciones y tecnologías relacionadas que apoyarán y permitirán comprender las contribuciones de esta investigación. De esta forma, en primer lugar, en la sección 3.1 se incluyen las definiciones y terminología utilizada en el documento, lo cual permitirá proyectar un mapa conceptual de la tesis doctoral. Seguidamente, la sección 3.2 describe el proceso de modelado de usuario y la sección 3.3 incluye los mecanismos para la recuperación de datos personales almacenados en distintas entidades. Las secciones 3.4 y 3.5 abordan la descripción de las áreas de gestión de identidad y privacidad respectivamente. Finalmente, la sección 3.6 presenta una descripción de los modelos de negocio existentes en el ecosistema de datos actual.

3.1 Definiciones y terminología

A continuación, se establecen las definiciones y terminología relacionadas con la tesis doctoral que se utilizará a lo largo de todo el documento. En primer lugar, se conceptualizan los términos datos personales, perfil de usuario, modelo de usuario e identidad, indicando su definición, las distintas tipologías, usos y áreas de investigación relacionadas. Todo ello permitirá establecer, en último lugar, un mapa conceptual en torno a los datos personales de los usuarios que refleja las relaciones entre los conceptos presentados y las áreas de investigación que aborda la tesis doctoral. Cada una de las áreas señaladas se detalla individualmente en secciones posteriores de este capítulo.

3.1.1 Datos personales

Para esta tesis doctoral tomaremos la definición general de **datos personales** proporcionada por el Foro Económico Mundial que denota a éstos como los datos digitales creados por y sobre un individuo o persona (WEF, 2011). Esto incluye no sólo los datos suministrados directamente por los usuarios cuando, por ejemplo, se registran en un servicio web, sino también aquellos datos que los proveedores de servicios pueden capturar o crear sobre sus usuarios.

La anterior definición da lugar a una categorización de los datos personales basada en la manera en la que los datos se originan. De esta forma, se identifican cuatro categorías principales (OECD, 2014):

- **Datos proporcionados:** son datos producidos explícitamente por los usuarios a través de distintas acciones como, por ejemplo, completar un formulario de registro para darse de alta en un servicio, compartir un pensamiento o impresión en una red social, rellenar los datos necesarios para realizar una transacción bancaria, etc. Aquí, los usuarios son conscientes de la realización de dichas acciones y el hecho de que están proporcionando estos datos a terceras entidades es obvio o al menos, intuitivo.
- **Datos observados:** son datos que terceras entidades, generalmente proveedores de servicios, capturan y almacenan cuando el usuario usa e interactúa con distintos servicios sin que éste sea completamente consciente de dichas acciones. Ejemplos de este tipo de datos son: aquellos obtenidos de *cookies*, historiales de navegación o sensores, como, por ejemplo, datos de localización, tiempos de visita por página, acciones realizadas con el cursor en una web, etc.
- **Datos derivados:** son datos generados por los proveedores de servicios a partir de otros datos basándose en mecanismos matemáticos básicos o razonamientos simples que pueden utilizarse con fines predictivos en otros procesamientos. Por ejemplo, la edad de un usuario obtenida a partir de su año de nacimiento, la rentabilidad de un usuario teniendo en cuenta las páginas que visita y las compras que realiza, etc. Generalmente, el usuario desconoce que los proveedores producen estos datos.
- **Datos inferidos:** son datos avanzados obtenidos por los proveedores de servicios a través del análisis probabilístico y/o la detección de correlaciones entre otros datos con el fin de predecir distintas características o comportamientos de los usuarios. Como en los dos casos anteriores, los usuarios no participan en la creación de estos datos y desconocen sus valores asociados, aunque éstos pueden percibir distintos comportamientos de los proveedores de servicios que hacen entrever la existencia de dichos datos. Ejemplos de datos inferidos son: la valoración de riesgo de crédito, la esperanza de vida o la probabilidad de realizar una compra, entre otros.

El WEF (2011) propone una clasificación similar a la anterior aunque no hace distinción de las dos últimas categorías y las denota de forma global como datos inferidos.

Además de la categorización anterior, podemos encontrar muchas otras que consideran aspectos particulares de los datos personales. A modo de ejemplo se citan las siguientes (OECD, 2014), (Asif, 2014), (*Future of Identity in the Information Society* – FIDIS, 2005) y (Barla, 2010):

- **Nivel de sensibilidad.** En distintos contextos, se identifican como datos sensibles a los datos más íntimos de los usuarios cuya utilización inadecuada producen un mayor impacto sobre la persona a la que aplica. Por ejemplo, distintas legislaciones (entre ellas la Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal) incluyen como datos sensibles a aquellos que denotan ideologías políticas, creencias, religión, afiliación sindical, salud, origen racial o vida sexual de las personas.
- **Grado de identificabilidad.** Éste indica cómo puede influir un dato en identificar a una persona determinada.
- **Temporalidad.** Pueden encontrarse distintas clasificaciones de los datos personales que tienen en cuenta sus características temporales. Así, tal y como señala (Barla, 2010), los datos personales se pueden clasificar de acuerdo a su validez en el tiempo como datos a **corto o largo plazo**. Los primeros reflejan características actuales y presentes de los usuarios que pueden variar en un corto espacio de tiempo, como por ejemplo, el contexto en el que se encuentra una persona. Por otra parte, los datos a largo plazo representan conocimiento más estático de los usuarios, como información de contacto o preferencias, que se mantienen durante un periodo de tiempo prolongado.

De manera más específica, tomando la clasificación de (FIDIS, 2005), los datos personales pueden clasificarse de acuerdo a su temporalidad como:

- **Permanentes dados:** representan características de las personas sobre las que no tienen influencia. Por ejemplo, características biológicas como el color de ojos, sexo, etc.
 - **Permanentes adquiridos:** son características que las personas adquieren bajo determinadas circunstancias o por la realización de una acción deliberada, como la calificación de un examen, comportamientos en determinados dominios, etc.
 - **Persistentes:** indican características de los usuarios que no son permanentes pero que pueden mantenerse a largo plazo, por ejemplo, varios años. La dirección de contacto física de una persona, el puesto de trabajo o el estado civil, entre otros, son ejemplos de este tipo de datos.
 - **Temporales:** incluye características dinámicas de los usuarios que varían a corto plazo, como por ejemplo, la localización o humor de una persona en un determinado momento.
- **Ámbito** de las personas a las que aplica. Por ejemplo, datos de empleados, ciudadanos, minorías, etc.

- **Dominio** en el que aplica, como por ejemplo, trabajo, educación, ocio, gobierno, salud, justicia, etc.
- **Independencia del dominio**, que determina si un dato personal es aplicable únicamente en un dominio particular (por ejemplo, preferencias e intereses sobre un elemento concreto) o si por el contrario, contiene características de los usuarios que pueden aplicarse en otros dominios (tales como la edad, características personales, etc.).
- **Propósito** para el que va ser usado. Por ejemplo, para uso personal, con fines científicos, por motivos económicos o legales.

Independientemente de las categorizaciones anteriores que determinan aspectos particulares de los datos personales, éstos pueden agruparse de acuerdo a su **naturaleza** o las características personales que describen de un usuario. En sentido, pueden encontrarse en la literatura multitud de clasificaciones con mayor o menor granularidad. A modo de ejemplo, dentro del proyecto europeo FIDIS (2005) se identificaron diez categorías principales, que han sido ampliadas para incluir otras características y ejemplos dentro de las mismas:

1. **Información de identificación:** son datos que se utilizan para referirse a una persona, por ejemplo, su nombre, apodo, apellidos o cualquier identificador dentro de un dominio, como el número de seguridad social, carné de conducir, etc.
2. **Características biológicas:** describen aspectos físicos de las personas, incluyendo información fisiológica (como la el peso o la altura), características relacionadas con la salud (enfermedades o resultados de análisis clínicos), identificadores biométricos (como la huella dactilar), etc.
3. **Localización:** determina la posición geográfica de una persona. Dependiendo de su temporalidad, ésta puede indicar la dirección física de una persona o la posición actual en la que se encuentra.
4. **Activos y transacciones:** refiere al conjunto de bienes, productos o servicios, que posee o puede adquirir una persona. Ejemplo de este tipo de datos son: el salario, las propiedades físicas, productos adquiridos, por ejemplo, en una tienda digital, etc.
5. **Ciudadanía y justicia:** incluye información de una persona desde una perspectiva político-legal, como, los antecedentes penales, partidos políticos a los que está afiliados, etc.
6. **Preferencias:** engloba el conjunto de intereses y elementos favoritos de una persona dentro de un dominio. Por ejemplo, los colores y diseños preferidos en una interfaz, las películas favoritas, etc.
7. **Competencias:** describen las capacidades de una persona, tanto certificadas por una entidad competente (por ejemplo, el carné de conducir o un título universitario), como aquellas que definen sus particularidades o conocimientos en un dominio.

8. **Funciones, posiciones, y roles:** indica el lugar que ocupa una persona o el rol que asume dentro de un dominio u organización, así como las tareas que realiza. Por ejemplo, puesto dentro de una empresa (director, empleado, etc.), rol dentro de un árbol genealógico (madre, hijo, tío, etc.).
9. **Características sociales:** incluye todo tipo de información de una persona como parte de un grupo. Por ejemplo, relaciones que mantiene, afiliaciones, reputación, rol dentro de una comunidad, etc.
10. **Características psicológicas:** describen los comportamientos y actuaciones de las personas, como su motivación, estilo cognitivo, personalidad, etc.

Gráficamente, la siguiente figura representa otra taxonomía con más detalle que la anterior donde se pone en evidencia el universo de datos personales que pueden asociarse a una persona.

Figura 3-1. Taxonomía de datos personales (Halim, 2011).

3.1.2 Perfil y modelo de usuario

En el área de Interacción persona-computador (en inglés, *Human-computer interaction* – HCI) y aquellas relacionadas con los sistemas adaptativos que aspiran a ofrecer a los usuarios mejores experiencias de uso y con ello, a adecuar su respuesta a las características particulares de los usuarios, los datos personales se engloban bajo los términos perfil y modelo de usuario.

La diferenciación entre perfil y modelo de usuario no está claramente establecida en la literatura y a menudo, ambos términos se usan como sinónimos o uno único se aplica con significado de ambos. Tomando la diferenciación que hacen algunos autores (Fröschl, 2005), el

término **perfil de usuario** refiere al conjunto de datos personales que se almacenan sin ninguna descripción o interpretación de los mismos. Dependiendo de las necesidades de negocio y de la cantidad de datos personales contenidos en los perfiles de usuario, los proveedores de servicios pueden analizar e interpretar dichos perfiles para elaborar distintos modelos de usuario. Así, un **modelo de usuario** es una representación de lo que un proveedor de servicios entiende que es un usuario específico en un contexto determinado. En otras palabras, un modelo de usuario es un conjunto de datos personales que reflejan una apreciación de lo que es, prefiere, quiere o va realizar un usuario, entre otros, y que permite a un proveedor de servicios responder de forma individualizada ante sus usuarios.

Tradicionalmente, los modelos de usuario se han basado en la representación de cinco características principales de los usuarios (Brusilovsky y Millán, 2007): **conocimiento** (especialmente relevante en áreas de aprendizaje electrónico), **intereses y preferencias** (ampliamente utilizados en sistemas recomendadores, buscadores, etc.), **objetivos y tareas** (para representar los propósitos inmediatos de los usuarios y anticiparse a sus necesidades), **antecedentes e historial** (que representa las aptitudes y destrezas del usuario en otros dominios) y **características individuales** (por ejemplo, el estilo de aprendizaje o tipo de personalidad). De esta forma, los modelos de usuario pueden utilizarse para, por ejemplo, recomendar nuevos contenidos o productos, personalizar la interacción o los resultados de un servicio, anticiparse a las necesidades de los usuarios, clasificar a un usuario dentro de un grupo determinado, etc. (Frías-Martínez, et al., 2005). En definitiva, de manera general, un modelo de usuario se puede destinar para mejorar la experiencia de uso de un servicio pero en muchas ocasiones, éste puede ser la pieza clave para ofrecerlo. Por ejemplo, en el contexto de un servicio de recomendación de productos financieros que ofrece a sus usuarios distintas posibilidades de inversión en función de sus características personales, es esencial para el éxito del servicio la existencia de modelos de usuario adecuados que reflejen las preferencias y necesidades particulares de los usuarios. En cambio, la utilización de estos mismos modelos por parte de una entidad bancaria puede suponer un valor extra añadido a los servicios principales de la entidad.

Dados los múltiples objetivos por los que los perfiles y modelos de usuario son creados, éstos pueden contener datos personales muy dispares. En términos generales, un perfil de usuario puede contener datos personales proporcionados y/o observados como, por ejemplo, datos demográficos, historiales de navegación, registro de transacciones, etc. En consecuencia, cualquier dato personal que no haya sido previamente procesado. Por otra parte, los modelos de usuario se centran en la representación y generación de conocimiento del usuario y por lo tanto, incluyen principalmente datos personales derivados e inferidos, aunque también pueden incluir únicamente datos proporcionados y observados que representen características relevantes de los usuarios y que pueden utilizarse para particularizar la respuesta de un sistema o servicio determinado.

Un usuario puede tener asociados múltiples perfiles y modelos de usuario en distintos contextos e incluso, en un dominio específico, éste puede tener varios perfiles y/o modelos de usuario. Esto último depende de cómo el proveedor de servicios de un dominio determinado organice y estructure los datos personales de los usuarios en sus sistemas. Tomando como ejemplo una compañía bancaria, aunque lo normal es que existan varios perfiles de usuario

dentro de una organización tan compleja, si simplificamos su estructura, podría considerarse un único perfil de usuario que contuviera los datos demográficos del cliente, sus productos contratados, sus transacciones financieras y sus últimos anuncios y páginas consultadas. De este perfil podrían obtenerse distintos modelos de usuarios que indiquen la adecuación o los intereses del cliente sobre distintos productos financieros, la capacidad de solvencia ante un crédito, la liquidez para una compra, etc.

Con todo lo anterior queda patente que un modelo de usuario puede entenderse de muy distintas formas. Con el objeto de conceptualizar sus dimensiones principales, y tomando como referencia a (Asif, 2014), puede establecerse que el diseño e implementación de un modelo de usuario es la combinación de soluciones específicas dentro de las siguientes cuatro perspectivas (ver Figura 3-2):

- **De simples a complejos.** Los modelos de usuario pueden ser entendidos como elementos simples, perfiles de usuario, que engloban datos personales básicos de los usuarios (por ejemplo, información demográfica o preferencias indicadas a través de un formulario o sistemas de votaciones) o más complejos cuando consideran datos personales inferidos o derivados y requiere el procesamiento y análisis de distintos datos personales.
- **De corto a largo plazo.** Los modelos de usuario pueden clasificarse en función de su tiempo de vida o la validez del modelo dentro de un contexto determinado. Así, dependiendo de las necesidades específicas, se pueden diseñar modelos de usuario con un corto plazo de tiempo asociado, por ejemplo, aquellos relacionados con la ubicación de un usuario o una sesión de navegación. Por otra parte, puede haber modelos de usuario de más largo plazo que consideran aspectos más estáticos de los usuarios, por ejemplo, sus preferencias audiovisuales, vivienda habitual, etc.
- **De estereotipos a altamente adaptativos.** Los modelos de usuario se pueden crear considerando categorías genéricas de usuarios (estereotipos) o características individuales de los mismos. De esta forma, habrá sistemas que clasifiquen a un determinado usuario dentro de un grupo preestablecido de estereotipos y otros, que implementen mecanismos para representar al usuario individualmente y por consiguiente, puedan responder con una mayor adaptabilidad.
- **De estáticos a dinámicos.** Los modelos de usuario pueden evolucionar en el tiempo y por ello, se pueden diseñar siguiendo un enfoque estático o dinámico. Un modelo de usuario estático no evoluciona y mantiene los valores con los que se creó a lo largo del tiempo. Sin embargo, en el otro extremo, un modelo de usuario dinámico evoluciona a medida que lo hace el usuario, y con ello, los datos personales relacionados con éste, lo cual permite adaptarse mejor a la realidad.

Figura 3-2. Tipos de modelos de usuario (Asif, 2014).

La obtención de modelos de usuario requiere llevar a cabo un proceso, denominado **modelado de usuario**, que se detallará en el apartado 3.2 Proceso de modelado de usuario. Únicamente, destacar que dicho proceso depende principalmente de los enfoques o soluciones consideradas dentro de cada una de las perspectivas anteriores. En términos generales, el modelado de usuario engloba distintas etapas relacionadas con la recolección, la integración, el procesamiento, el análisis y el uso de datos personales.

3.1.3 Identidad digital

Para la definición de identidad y sus términos asociados, nos apoyaremos en organismos de estandarización como la Unión Internacional de Telecomunicaciones (del inglés, *International Telecommunication Union* – ITU) o la Organización Internacional de Normalización (del inglés, *International Organization for Standardization* – ISO), cuyos objetivos son la estandarización en el contexto de las telecomunicaciones o la industria a nivel internacional.

Basándonos en la definición del estándar ITU-T X.1250 (ITU, 2009), podemos definir la **identidad** digital de una persona como la representación de dicha persona en forma de uno o más elementos de información (datos personales) que permiten distinguir a ésta dentro de un contexto determinado (generalmente en un proveedor de servicios). Cada persona puede estar representada por una **identidad holística** que abarca todos los datos personales que caracterizan a tal persona, sin embargo, frecuentemente, dicha representación holística no va más allá de la teoría, puesto que el número de posibles datos personales de una identidad es indefinido. Desde una perspectiva más pragmática, al igual que ocurre con los perfiles y modelos de usuario asociados a una persona específica, una persona tiene asociadas una o más identidades en distintos contextos y en un dominio específico puede tener varias identidades. Por ejemplo, si una persona trabaja y al mismo tiempo, es cliente de un banco, ésta tendrá dos identidades frente a dicho banco: una asociada a sus características como trabajadora y otra como cliente del banco.

Al concepto de identidad se asocian frecuentemente los términos: atributo e identificador. Un **atributo** es un dato personal o conjunto de éstos que especifica una característica de una

persona. Por otra parte, un **identificador** es uno o más atributos que se utilizan para distinguir (identificar) a una persona dentro de un contexto determinado. Así, cada identidad se puede representar con al menos uno o más atributos y/o identificadores. La ISO, concretamente en su estándar ISO/IEC 29100 (ISO, 2011), especifica definiciones similares a las anteriores pero designa **información personalmente identificable** (del inglés, *Personally Identifiable Information* – PII), como cualquier dato o conjunto de éstos que bien puede utilizarse para identificar a la persona a la que aplica, o bien puede vincularse directamente o indirectamente con una persona.

Una identidad puede contener entre sus atributos e identificadores cualquiera de los tipos de datos personales mencionados (proporcionados, observados, derivados o inferidos) y normalmente está asociada a una **credencial** o conjunto de datos que certifican la evidencia de la identidad o parte de ésta y unos derechos declarados. Principalmente, existen dos tipos de credenciales: unas sirven para autenticar una identidad declarada y así, aumentar el grado de confianza de los atributos incluidos, y otras certifican unos derechos declarados, como, por ejemplo, la autorización a un recurso o cualquier otro privilegio. Ejemplos de credenciales asociados con una identidad son contraseñas, parámetros biológicos, certificados digitales, etc. Por otra parte, para preservar el anonimato de una identidad, ésta puede tener asociado distintos **pseudónimos** que permite identificar a una persona únicamente a las entidades que los asignan.

El término identidad tradicionalmente se aplica dentro de las áreas de **Gestión de Identidad y Privacidad** para llevar a cabo funciones relacionadas con el control de acceso, autorización, autenticación, descubrimiento, seguridad, etc. El término PII es propio del área de Privacidad. En las secciones 3.4 Gestión de identidad y 3.5 Privacidad se detalla el estado del arte relacionado con dichas áreas.

3.1.4 Mapa conceptual de la tesis doctoral

Como se puede deducir de las definiciones anteriores, los proveedores de servicios se apoyan en el uso de los datos personales para llevar a cabo distintas funcionalidades dentro de sus servicios. Dependiendo del uso que se dé a dichos datos y con ello, del ámbito de aplicación considerado, éstos se denominan como identidad, atributo, identificador, credencial, información personalmente identificable, perfil de usuario, modelo de usuario, entre otros. Aunque todos estos términos están íntimamente relacionados y un mismo dato puede corresponder a más de uno de los términos mencionados, existen diferencias sutiles entre los mismos.

La Figura 3-3 representa el mapa conceptual de la tesis doctoral donde se engloban los conceptos descritos anteriormente, así como las áreas de investigación relacionadas. En la parte central, se observan los distintos tipos de datos personales definidos en el apartado 3.1.1 (proporcionados, observados, derivados e inferidos) y como éstos pueden corresponder, a su vez, con distintas identidades digitales, perfiles de usuario y modelos de usuario.

Figura 3-3. Mapa conceptual de la tesis doctoral (San Miguel, et al., 2014c).

Simplificando la complejidad de cada una de las áreas de investigación abordadas, se puede establecer que, por una parte, los proveedores de servicios se apoyan en el uso de datos personales de los usuarios para construir identidades digitales que facilitan el acceso y gestión los recursos personales (en la Sección 3.4 Gestión de identidad se describe en profundidad los procesos relacionados). Por otra parte, los proveedores de servicios pueden llevar a cabo un proceso de modelado de usuario (descrito en la Sección 3.2 Proceso de modelado de usuario) que permite obtener conocimiento de los usuarios para responder de forma individualizada a las características particulares de éstos. De forma transversal a cada una de la áreas indicadas, el área de investigación de privacidad (ver Sección 3.5 Privacidad) aborda aspectos relacionados con la utilización de los datos personales de los usuarios y el impacto de esto sobre las personas.

Finalmente, cabe señalar que denominaremos de forma general a las personas que hacen uso de los servicios ofrecidos por proveedores del sector público o privado como usuarios, personas o clientes indistintamente, aunque existen otros términos más específicos en otras áreas de investigación como, por ejemplo, sujeto de datos o principal en el área de privacidad.

En las siguientes secciones se presenta el estado del arte de las áreas de investigación mencionadas.

3.2 Proceso de modelado de usuario

Tradicionalmente, el modelado de usuario es un proceso unilateral a través del cual un proveedor de servicios genera modelos de usuario que cubren sus necesidades de negocio en un dominio específico. El proceso de modelado de usuario depende directamente del tipo de modelo considerado que, tal y como se describió en 3.1.2 Perfil y modelo de usuario, viene determinado por la combinación de soluciones específicas dentro de cuatro perspectivas que

definen modelos de usuario que van de simples a complejos, de corto a largo plazo, de estereotipo a altamente adaptativos y de estáticos a dinámicos.

En términos generales, se puede establecer que el modelado de usuario abarca tres etapas fundamentales (Brusilovsky, 1996): recolección de datos personales, generación o inferencia y aplicación del modelo de usuario generado y éstas pueden sucederse cíclicamente.

La Figura 3-4 representa las etapas principales del modelado de usuario que lleva a cabo un proveedor de servicios y su relación con el usuario. En primer lugar, el proveedor de servicios realiza la recolección de los datos personales que le permitirán producir modelos de usuario. Seguidamente, en la etapa de generación produce conocimiento de los usuarios en forma de modelos de usuario que, finalmente, aplica con distintos fines. Las tres etapas anteriores pueden sucederse repetidamente si se considera un proceso de realimentación que permite capturar la reacción del usuario ante el modelo aplicado y utilizar ésta en la etapa de recolección.

Figura 3-4. Etapas principales del modelado de usuario (Barla, 2010) y (San Miguel, et al., 2015c).

A continuación, se detallan los aspectos y técnicas más relevantes para cada una de las etapas identificadas. En último lugar, se incluye un apartado con otros trabajos relacionados del área de modelado de usuario.

3.2.1 Recolección de datos personales

La recolección de datos personales es la primera etapa del proceso de modelado de usuario que permite obtener los datos que posteriormente se utilizarán para generar modelos de usuario. Existen dos métodos principales para llevar a cabo la recolección de datos personales: explícito e implícito, aunque pueden aplicarse enfoques híbridos combinando los dos métodos anteriores (Asif, 2014).

3.2.1.1 Método explícito

En el método explícito los datos personales se obtienen cuando los usuarios los introducen o aportan directamente en el servicio. El mecanismo clásico de este enfoque son formularios que los usuarios rellenan con sus datos personales, por ejemplo, un formulario de registro, una encuesta de satisfacción u otros sistemas de valoración de contenidos. De esta manera, los datos obtenidos por este método corresponden a la categoría de **datos proporcionados** descritos en 3.1.1 Datos personales.

La mayor ventaja asociada a este método es que permite recolectar los datos personales de los usuarios rápidamente, sin embargo, depende directamente de la disposición de los usuarios para realizar las tareas requeridas, por ejemplo, rellenas los formularios provistos. Además, aunque pueden implementarse distintos mecanismos para que los usuarios actualicen los datos introducidos (por ejemplo, un sistema de notificaciones), en general, los datos proporcionados tienen característica de estáticos, puesto que una vez que el usuario los introduce éstos no suelen cambiar a lo largo del tiempo.

3.2.1.2 Método implícito

El método implícito consiste en la aplicación de técnicas que permiten la recolección de datos personales de los usuarios cuando éstos interactúan con un determinado servicio. A diferencia del método anterior, el usuario no tiene que realizar acciones específicas para aportar los datos, únicamente utilizar el servicio o sistema considerado. Por ejemplo, se pueden obtener datos personales de los registros de navegación, cookies, historiales de búsqueda o del contexto (como el dispositivo, navegador o localización) sin que el usuario perciba dicha recogida, aunque haya tenido que aprobarla o autorizarla previamente o al menos, haya sido informado correctamente de la aplicación de estos métodos.

La siguiente figura muestra un ejemplo de aviso (conforme al Real Decreto-ley 13/2012) que un usuario se puede encontrar cuando navega en un proveedor de servicios que aplica cookies como método implícito de recolección de datos. Éstas permiten obtener datos de los usuarios que pueden utilizarse para mejorar la experiencia de uso del servicio y únicamente requiere que el usuario navegue por la web del proveedor.

Figura 3-5. Ejemplo de aviso sobre la utilización de cookies⁶.

⁶ El Corte Inglés, <http://www.elcorteingles.es/>

Los datos obtenidos por este método corresponden a la categoría de **datos observados** descritos en 3.1.1 Datos personales. Generalmente, estos datos no están estructurados y tienen asociado un mayor esfuerzo para la siguiente etapa del procesamiento de modelado de usuario.

3.2.2 Generación de modelos de usuario

Una vez que los datos personales de los usuarios se han recolectado a través de los métodos anteriores, éstos se adaptan y almacenan generalmente en uno o varios perfiles de usuario, aunque también pueden almacenarse directamente como modelos de usuario, dependiendo de las estrategias específicas del proveedor de servicio que lleva a cabo el proceso.

Tal y como se expuso en 3.1.2 Perfil y modelo de usuario, un modelo de usuario puede estar formado por cualquiera de los tipos de datos personales descritos en 3.1.1 Datos personales (proporcionados, observados, derivados y/o inferidos), sin embargo, cuanto más complejo es un modelo de usuario, éste incluye mayor número de **datos derivados y/o inferidos**. En términos generales, la etapa de generación engloba todas las acciones que deben realizarse para producir modelos de usuario a partir de los datos personales obtenidos en la etapa anterior y esto incluye el procesamiento de los datos proporcionados y/u observados para crear y producir datos derivados y/o inferidos.

A continuación, se describen los mecanismos de generación más utilizados que, a su vez, determinan la representación de dichos modelos (Montaner, et al., 2003), (Gauch, et al., 2007) y (Barla, 2010): vectores de características, estereotipos, modelos basados en sistemas clasificadores y modelos basados en tecnologías semánticas. Finalmente, se incluye un apartado que resume y analiza las características de éstos.

3.2.2.1 Vectores de características

En este enfoque se generan modelos de usuario y con ello, se representan los usuarios, como un conjunto de pares elemento-valor que conforman un vector. Cada elemento simboliza un concepto específico del dominio o, menos frecuentemente, una característica del usuario, como, por ejemplo, productos financieros de un banco, enlaces de un sitio web, las preferencias musicales o las capacidades para realizar una tarea, y tiene asociado un valor que indica la actitud del usuario hacia dicho elemento (Gauch, et al., 2007). Para ilustrar esto, en un contexto bancario, a cada producto financiero se le puede asignar un valor binario que indique si el usuario lo ha contratado o no, o cada enlace de una página web puede tener asociado un valor numérico que indique las visitas del usuario para determinar, posteriormente, su interés.

Los vectores de características son muy sencillos de implementar y se utilizan en la mayoría de sistemas de recuperación de información, como buscadores, catálogos de películas o libros, tiendas electrónicas, etc. (Brusilovsky y Millán, 2007). Principalmente, para la generación de los vectores de características deben identificarse los elementos que conforman el vector y los algoritmos que permitan actualizar los valores asociados a cada elemento.

Existen otros enfoques similares a los vectores de características como, por ejemplo, los basados en palabras o también denominados **sacos de palabras**, que representan a los usuarios como un conjunto de términos, los **basados en contenido**, similares a las palabras pero considerando otro tipo de contenidos o las **matrices de características**, que consideran los datos históricos de un usuario o del resto de usuarios del sistema respecto a un determinado concepto (Gauch, et al., 2007).

3.2.2.2 Estereotipos

Los estereotipos definen modelos de usuario genéricos que permiten identificar y clasificar a los usuarios de un servicio dentro de éstos, típicamente, en función de sus características demográficas aportadas a través de un formulario de registro. Cada usuario que pertenece a un mismo estereotipo es tratado como el resto de usuarios de dicho grupo y en general, no se tienen en cuenta sus características individuales, aunque existen variaciones que permiten evolucionar los estereotipos con las particularidades de los usuarios.

La gran ventaja asociada a esta técnica es que con pocos datos personales se pueden suponer e inferir muchas características y actitudes de los usuarios. Sin embargo, la identificación y definición de los estereotipos requiere un gran esfuerzo inicial, lo cual requiere de conocimientos a nivel psicológico y estadístico de los usuarios y sus comportamientos. Por otra parte, la asignación de los usuarios a un estereotipo determinado puede generar decisiones incorrectas que se traducen en particularizaciones incorrectas del servicio y con ello, frustración de los usuarios.

Los mecanismos y técnicas asociados al modelado de usuario basado de estereotipos incluyen algoritmos de agrupamiento (en inglés, *clustering*) que se pueden aplicar para identificar patrones de usuarios dado un conjunto de datos, técnicas de aprendizaje de máquinas o razonamiento para actualizar el estereotipo asociado con un usuario, etc. (Kay, 1994).

3.2.2.3 Modelos basados en clasificadores

Los sistemas clasificadores (Montaner, et al., 2003) generan modelos de usuario basándose en distintos algoritmos que asignan a un elemento de entrada una respuesta determinada.

Generalmente, los algoritmos utilizados por los sistemas clasificadores tienen su origen en técnicas de aprendizaje de máquinas como redes Bayesianas, árboles de decisiones, reglas inductivas o redes neuronales artificiales, que permiten la reestructuración, adaptación y predicción de las respuestas con su entrenamiento. Aquí, los modelos de usuario se mantienen y representan con el propio algoritmo o estructura de aprendizaje utilizada por el sistema clasificador.

La generación de estos sistemas está muy enfocada a resolver tareas concretas y precisas de un sistema y requieren un gran esfuerzo de entrenamiento para su óptima utilización.

3.2.2.4 Modelos basados en tecnologías semánticas

Dentro de las tecnologías semánticas, las ontologías juegan un papel fundamental en el proceso de modelado de usuario, puesto que éstas pueden utilizarse en cada una de las etapas

de dicho proceso, tal y como señalan Sosnovsky y Dicheva (2010) que describen los distintos usos que se les ha dado en el área a lo largo del tiempo.

Una **ontología** es un artefacto de ingeniería formado por un vocabulario que representa una realidad concreta y una serie de supuestos explícitos sobre el significado y la relación de dicho vocabulario (Guarino, 1998). Ésta puede utilizarse dentro de la etapa de generación del modelado de usuario tanto para modelar un dominio específico, como a los usuarios propiamente dichos, puesto que establece una red de términos enriquecidos con relaciones entre sí. Habitualmente, dichas relaciones corresponden a lógica de primer orden, lo cual permite deducir y concluir características de los usuarios a través de la definición de reglas de inferencia.

Existen distintas metodologías y lenguajes para definir ontologías como las indicadas por Yu (2011). La mayoría de las metodologías propuestas se sustentan en la reutilización de ontologías y vocabularios ya existentes, puesto que esto permite, a su vez, la reutilización y aprovechamiento de nuevo conocimiento generado.

En lo que respecta a ontologías específicas que aborden la presentación de datos personales, a modo de ejemplo, cabe mencionar la ontología *Friend of a Friend* (FOAF) (Brickley y Miller, 2014) que permite representar a las personas, sus actividades y relaciones en el contexto web y que ha sido extendida en distintas ocasiones para incluir nuevo conocimiento de las personas. Además, el *Web Ontology Language* (OWL) (W3C, 2004), recomendación del W3C, permite la definición de nuevas ontologías.

Conjuntamente, existen distintos lenguajes y herramientas que permiten extraer y deducir información almacenada semánticamente. Por ejemplo, para datos expresados en *Resource Description Framework* (RDF), que modelan información bajo un enfoque de tripletas, el *SPARQL Protocol and RDF Query Language* (SPARQL) (W3C, 2015), recomendación también del W3C, permite realizar consultas y búsquedas sobre dichos contenidos.

Además, existen distintas herramientas de inferencia, como son los razonadores, que pueden procesar el conocimiento almacenado semánticamente para detectar inconsistencias, deducir e inferir nuevos hechos y consecuencias (Yu, 2011). Se pueden encontrar razonadores integrados en la mayoría de entornos de desarrollo para aplicaciones semánticas, como por ejemplo Jena⁷ o Protégé⁸, o también existen razonadores externos como Pellet⁹.

3.2.2.5 Resumen de los mecanismos de generación

En este apartado se han presentado cuatro mecanismos principales que permiten llevar a cabo la generación de modelos de usuario: vectores de características, estereotipos, modelos basados en sistemas clasificadores y modelos basados en tecnologías semánticas. La elección o combinación de éstos para su aplicación dependerá de las características de los modelos de usuario deseados, así como del uso que se los vaya a dar.

⁷ Apache Jena, <http://jena.apache.org/>

⁸ Protégé, <http://protege.stanford.edu/>

⁹ Pellet, <https://github.com/complexible/pellet>

Los tres primeros mecanismos descritos generalmente están enfocados en soluciones especializadas en un dominio de aplicación muy concreto y por lo tanto, tienen asociado una gran dificultad para permitir la compartición de los modelos de usuario entre proveedores de servicios y aún más cuando éstos pertenecen a dominios distintos. Esto es debido a que los modelos de usuario se mantienen y representan con una estructura y semántica que viene determinada por el dominio de aplicación, lo cual permite resolver tareas concretas y precisas en la aplicación de un modelo de usuario. No obstante, dichas estructuras son rígidas y dificulta su utilización en otros contextos distintos a los que fueron creados. En este sentido, la generación de modelos de usuarios basada en tecnologías semánticas trata de solventar los problemas de intercambio, comunicación e interoperabilidad mencionados. Más adelante, en Capítulo 5 (5.1.2 Especificación complementaria: requisitos no funcionales), se profundizará sobre este aspecto y la necesidad de soluciones interoperables en un contexto distribuido para la generación de modelos de usuarios. Asimismo, en el Capítulo 6 se presenta una solución basada en tecnologías semánticas para la generación y compartición de modelos de usuario.

3.2.3 Aplicación del modelo de usuario y realimentación

La última fase del modelado de usuario es el uso o la aplicación del modelo de usuario generado. Como se describió en 3.1.2 Perfil y modelo de usuario, en términos generales, un modelo de usuario puede utilizarse para mejorar las prestaciones de un servicio con características relacionadas con la personalización, recomendación o predicción, aunque éste también puede ser la pieza clave que permita ofrecer dicho servicio.

La aplicación de los modelos de usuario generados requiere llevar a cabo mecanismos que permitan identificar a los usuarios sobre los que refiere el modelo de usuario y establecer las situaciones y contextos concretos en los que aplica dicho modelo. Todo ello permite que los proveedores de servicios particularicen su respuesta en el lugar y momento preciso para un usuario concreto.

Cuando un proveedor de servicios aplica un modelo de usuario, el propio usuario al que aplica dicho modelo aprecia y percibe la particularización que se ha producido en el servicio ofrecido. Esta apreciación y reacción del usuario puede recogerse para reiniciar el proceso de modelado de usuario y así, permitir el refinamiento y enriquecimiento de los modelos de usuarios existentes.

Las reacciones de los usuarios se denominan formalmente en el área de modelado de usuario como **feedback**. Típicamente, éste es de dos tipos (Montaner et al., 2003): positivo (cuando se refleja una buena actitud del usuario, por ejemplo, que le gusta las recomendaciones ofrecidas) o negativo (cuando la aplicación el modelo ha sido ignorada o rechazada por el usuario).

De forma similar a la etapa de recolección de datos personales, la apreciación del usuario se puede monitorizar a través de mecanismos de realimentación explícitos, implícitos o híbridos, sin embargo, puede no llevarse a cabo, dependiendo de la implementación concreta que realice el proveedor de servicios (Montaner et al., 2003). Por ejemplo, un proveedor de servicios dedicado a la venta de películas puede aplicar modelos de usuario para recomendar

nuevas películas que se adapten a los intereses y preferencias de sus usuarios. Estas películas se mostrarían a sus usuarios y sobre cada una de ellas se podría incluir una escala de valoración para que el usuario indique directamente la adecuación de la recomendación (método explícito) o se podría deducir que la película es una buena recomendación porque el usuario ha seleccionado, consultado o comprado dicha película (método implícito). La implementación de ambos métodos por parte de un proveedor daría lugar a un método híbrido.

Generalmente, la realimentación dada por el feedback de los usuarios se aplica en servicios donde éstos son la pieza clave para ofrecerlo, puesto que requiere una evolución y adaptación de los modelos de usuario para mejorar la propia prestación del servicio.

3.2.4 Otros trabajos relacionados

En los apartados anteriores se ha presentado una visión genérica y global del proceso de modelado de usuario a través de la descripción de sus etapas fundamentales: recolección de datos, generación de modelos de usuario y aplicación de éstos. Los primeros trabajos relacionados con este área datan de la década de los setenta (Carbonell, 1970) y desde entonces se han llevado a cabo multitud de investigaciones y avances. A modo de ejemplo, las referencias previamente citadas (Montaner, et al., 2003), (Brusilovsky y Millán, 2007), (Gauch, et al., 2007) y (Barla, 2010) incluyen distintos análisis del proceso de modelado de usuario, sus etapas y las tecnologías asociadas.

Tal y como se señaló en la introducción de esta sección, habitualmente la puesta en práctica del proceso de modelado de usuario, así como las investigaciones asociadas, se realiza desde el punto de vista de un único proveedor de servicios que requiere su utilización. No obstante, con el tiempo han aparecido soluciones puntuales, como, por ejemplo, los Sistemas Genéricos de Modelado de Usuario o los Servidores de Modelado de Usuario (Kobsa, 2001), que se basan en la **externalización del proceso de modelado de usuario** a sistemas especializados.

Ejemplos de los sistemas mencionados puede encontrarse en (van der Sluijs y Houben, 2005) que propone un componente denominado como *Generic User Model Component* para facilitar el almacenamiento de modelos de usuario a distintas aplicaciones, en (Heckmann, 2005) donde se presenta un entorno para facilitar intercambio de modelos de usuario en contextos ubicuos o en (Kobsa y Fink, 2006) que describe un servidor de modelos de usuario basado en el *Lightweight DirectoryAccess Protocol* (LDAP). Para ampliar esta información, en (Kobsa, 2001), (Viviani, et al., 2010) y (Carmagnola, et al., 2011) se detallan más ejemplos de productos comerciales y soluciones experimentales relacionados con este enfoque.

En términos generales, las soluciones identificadas abordan uno o varios aspectos relacionados con la interoperabilidad y el intercambio de modelos de usuario como, por ejemplo, el enfoque arquitectónico adoptado, los protocolos de comunicación entre sistemas, la representación de los datos y modelos de usuario, la gestión de datos distribuidos o la privacidad. En (Carmagnola, et al., 2011) se presenta un análisis de las dimensiones consideradas por distintos sistemas propuestos en la literatura.

Esta tesis doctoral se basa en la externalización del proceso de modelado de usuario, sin embargo, a diferencia de las soluciones detectadas, ésta se lleva a cabo mediante la distribución del proceso y la consideración de la participación de otras entidades como los usuarios y distintos proveedores de servicios. En concreto, en el Capítulo 6 se definirá el **proceso de modelado de usuario distribuido** como la realización de dicho proceso aprovechando los recursos que distintas entidades (fuentes de datos) ofrecen para generar modelos de usuario enriquecidos que cubren las necesidades específicas de proveedores de servicios (consumidores), considerando la participación del usuario y el control sobre sus recursos personales (datos y modelos de usuario).

En particular, lo anterior requiere la utilización de distintos mecanismos de recuperación de datos como los que se describirán en la sección 3.3 y el establecimiento de una infraestructura de gestión de identidad y privacidad que se apoya en tecnologías como las descritas en 3.4 y 3.5 respectivamente. A continuación, se presenta el estado del arte relacionado con estos aspectos, más adelante, en los capítulos 6, 7 y 8 las contribuciones relacionadas con el proceso de modelado de usuario distribuido que se realiza conforme a la arquitectura establecida en el capítulo 5.

3.3 Mecanismos de recuperación de datos personales almacenados en fuentes de datos externas

Como se describió en 3.2 Proceso de modelado de usuario, los proveedores de servicios pueden aplicar dos métodos principales (explícito e implícito) que permiten llevar a cabo la recolección de datos proporcionados y/u observados de los usuarios. Además, en la etapa de generación de modelos de usuarios, dichos datos pueden procesarse para producir datos derivados y/o inferidos que conforman distintos modelos de usuario.

Ahora bien, un proveedor de servicios puede ofrecer los datos personales que posee (independientemente de si éstos corresponden a datos proporcionados, observados, derivados y/o inferidos) a terceras entidades para que éstas los recuperen y los aprovechen con distintos fines de los que fueron creados. De esta forma, el proveedor de servicio se convierte en un proveedor de datos y puede denominarse como **fuentes de datos** desde la perspectiva de la tercera entidad que recupera sus recursos.

Existen dos tipos principales de fuentes de datos respecto al acceso a los datos que contienen: públicas o privadas, aunque una fuente puede catalogarse dentro de ambos grupos, dependiendo de los datos que se consideren. Las **fuentes públicas** de datos personales incluyen datos que son accesibles de forma equitativa para cualquier entidad de una red pública. Por ejemplo, un proveedor de servicios o los propios usuarios de Internet que acceden a contenido de un periódico en línea sin suscripción para indexar o leer una noticia de una persona determinada respectivamente. Por su parte, en las **fuentes privadas** de datos, el acceso a los datos personales sólo pueden realizarlo entidades autorizadas. Esta autorización puede venir del propio proveedor de servicios (generalmente motivado por fines económicos o acuerdos comerciales) o del usuario al que refieren los datos personales (por ejemplo, para dar acceso a un servicio de otro proveedor que requiere los datos almacenados). Como

ejemplo de una fuente de datos híbrida, cabe citar los servicios de redes sociales que albergan, por una parte, contenido generado por los usuarios al que únicamente pueden acceder los usuarios del servicio y por otra parte, perfiles básicos (por ejemplo, nombre, apellidos y foto de identificación) de acceso público.

Los **mecanismos para la recuperación** de datos personales almacenados en fuentes de datos externas dependen principalmente del tipo de fuente considerada (pública o privada) y de la propia fuente de datos y en concreto, de los servicios que ésta ofrezca para llevar a cabo la recuperación de sus recursos. En el contexto de esta investigación, se han identificado cuatro mecanismos principales que pueden llevarse a cabo a través de distintas tecnologías: programas araña, ficheros estáticos, peticiones y suscripciones. A continuación, se describe cada uno de ellos. Es importante aclarar, de nuevo, que para realizar la recuperación de recursos custodiados por fuentes de datos privadas es necesario previamente llegar a acuerdos comerciales entre las entidades involucradas o bien obtener el permiso del usuario, por ejemplo, a través de tecnologías de autorización delegada como las descritas en 3.4.2 Tecnologías para la delegación de autorización.

3.3.1 Programas araña

Los programas araña (conocidos por sus términos en inglés *spiders* o *crawlers*) son el mecanismo más utilizado para recuperación de datos de fuentes públicas. Éstos son aplicaciones software o robots inteligentes (denominados también con su aféresis *bots*) especializados en rastrear y descargar automáticamente información de Internet para su posterior tratamiento. Hoy en día son ampliamente utilizados por la mayoría de los motores de búsquedas.

Un programa araña parte de un conjunto inicial de direcciones web (denominadas como semillas) que va visitando para descargar su contenido, analizarlo y así, de acuerdo con unos criterios predefinidos, detectar datos relevantes que puede almacenar, por ejemplo, en una base de datos, o enlaces que suponen nuevas direcciones web a inspeccionar.

Existen cuatro **políticas** que definen el comportamiento de un programa araña y en consecuencia, determinan su diseño (Castillo, 2004): **de selección** (establece qué páginas rastrear o inspeccionar), **de frecuencia de visita** (indica cuándo se debe visitar una dirección concreta para mantener actualizada los datos recolectados), **de diplomacia** (determina los mecanismos a llevar a cabo para no saturar la fuente de datos consultada) y **de paralelización** (para coordinar la concurrencia de distintos procesos del programa).

La implementación de programas araña cubre tecnologías muy diversas relacionadas con el desarrollo web (*HyperText Markup Language* – HTML, JavaScript o *Asynchronous JavaScript And XML* – AJAX), la extracción de información (como los *parsers* que permiten analizar el texto de un contenido) o la indexación de contenidos que permite realizar consultas posteriormente (Cabrera, 2013). Además, existen distintos programas araña de código abierto como, por ejemplo, Crawler4j¹⁰ o Apache Nutch¹¹.

¹⁰ Crawler4j, <https://github.com/yasserg/crawler4j>

3.3.2 Ficheros estáticos de datos

El intercambio de datos personales de los usuarios a través de ficheros estáticos es uno de los mecanismos más sencillos y tradicionales para la recuperación de datos personales. Éste se basa en que la fuente de datos considerada pone a disposición de terceras entidades, incluyendo el propio usuario, un archivo que engloba los datos personales generados en dicha fuente.

Generalmente, los ficheros de datos engloban un perfil de usuario con información sin procesar y con formatos establecidos por la fuente de datos. Por ejemplo, Google (2015) ofrece a sus usuarios la posibilidad de exportar todos los datos generados en sus servicios descargando diversos ficheros de compresión (ZIP) que contienen distintos formatos dependiendo del tipo de datos considerados.

Las desventajas asociadas con los ficheros estáticos de datos personales es que éstos se pueden modificar fácilmente, por lo que se puede manipular el contenido de los mismos. Así, los mecanismos de seguridad asociados son muy débiles y aún más, cuando éstos requieren la intervención manual del usuario para la descargar del fichero. Asimismo, dicha intervención requiere un trabajo extra para los usuarios que no siempre están dispuestos a participar. Finalmente, los archivos recuperados tienen la característica de estáticos y para mantenerlos actualizados requeriría repetir el proceso de recuperación del fichero constantemente.

Con el objeto de solventar las desventajas e inconvenientes asociados con la recuperación a través ficheros estáticos y con ello, automatizar el proceso para reducir la participación de los usuarios al mínimo, existen otros mecanismos como los que se describen a continuación.

3.3.3 Mecanismos de petición

Un mecanismo de petición consiste en que una entidad externa a la fuente de datos considerada realiza una solicitud puntual de un recurso y obtiene dicho recurso en distintos formatos establecidos por la fuente de datos personales.

Las llamadas remotas a interfaces de programación de aplicaciones (del inglés, *Application Programming Interfaces* – API) se han convertido en el mecanismo de facto para permitir la reutilización de recursos, incluyendo dentro de éstos los datos personales, de servicios existentes. Tanto es así que, en los últimos años, ha surgido el término “economía API” (Gat y Succi, 2013) para referir al nuevo ecosistema en el que distintas compañías exponen sus recursos y activos de negocio a terceras entidades, con el objetivo de fomentar el descubrimiento de nuevas oportunidades y la creación de nuevos servicios, productos y en definitiva, activos económicos. Dada su relevancia, a continuación se detallan los aspectos y tecnologías más relevantes relacionados con la economía API.

¹¹ Apache Nutch, <http://nutch.apache.org/>

3.3.3.1 Recuperación de datos mediante interfaces de programación de aplicaciones

En la siguiente figura se ha representado una visión global del escenario sobre el que se sustenta el ecosistema de economía API. Aquí, se pueden observar distintos proveedores de APIs, sus usuarios, un consumidor de APIs que obtiene distintos recursos de los proveedores y el usuario final que disfruta de los servicios. A su vez, un consumidor de APIs podría también actuar como proveedor de APIs si ofrece sus recursos a terceras entidades.

Figura 3-6. Visión global del ecosistema de economía API (San Miguel, et al., 2015b).

Por consiguiente, se identifican tres **actores** principales dentro del ecosistema de economía API: proveedores de APIs, consumidor de APIs y el usuario final. Los primeros ofrecen y exponen los datos personales que custodian de sus usuarios u otros recursos que poseen a través de APIs a terceras entidades, denominadas como consumidores de APIs, para que creen nuevos servicios y activos que disfrutan los usuarios finales o personas físicas que, a su vez, también pueden hacer uso de los servicios ofrecidos por los proveedores de APIs.

Es importante destacar que en el caso particular del intercambio de datos personales a través de APIs, generalmente, se deberá obtener el consentimiento y la autorización del usuario para poderlo llevarlo a cabo, tal y como se detallará en 3.4.2 Tecnologías para la delegación de autorización.

El ecosistema de economía API crea oportunidades y beneficios a los tres actores involucrados (Gat y Succi, 2013). En primer lugar, los consumidores de APIs ven cómo se agiliza el proceso de creación y desarrollo de nuevos servicios, reduciendo tiempo y esfuerzo, al poder reutilizar recursos existentes. Además, entre dichos recursos también pueden encontrarse los propios usuarios de los proveedores de APIs y con ello, se constituye un nuevo modelo de distribución masivo capaz de aumentar la visibilidad, penetración y utilización de los nuevos servicios creados. Todo lo anterior fomenta la creación de nuevos activos que repercuten y aumentan el valor de los propios proveedores de APIs, ya que pueden beneficiarse de las nuevas innovaciones, descubrir nuevos usuarios, plataformas o dispositivos, extender su marca, etc.

Finalmente, respecto a los usuarios finales, éstos disfrutan de nuevos servicios que complementan, evolucionan y mejoran los servicios existentes.

El número de APIs disponibles ha crecido de forma exponencial en los últimos años y se espera que siga haciéndolo (Gat y Succi, 2013). Los ejemplos más característicos que han fomentado la nueva economía API son las plataformas de distribución de aplicaciones móviles, como por ejemplo App Store¹² y Google Play¹³, y los servicios de redes sociales, como Facebook y Twitter.

Desde un punto de vista técnico, las APIs suponen un conjunto de rutinas, protocolos y herramientas que permiten a desarrolladores externos acceder a distintas funcionalidades y recursos ofrecidos por un proveedor, abstrayendo de los detalles de implementación concretos. Su definición depende del proveedor de APIs considerado y por ello, no existe una especificación unificada para las APIs disponibles. Es decir, cada una contiene su propia descripción y la documentación asociada puede ser escasa o incluso, incorrecta, por lo que su utilización puede suponer un reto.

Tomando como ejemplo los servicios de redes sociales, existen distintas iniciativas cuyo objetivo es solventar los problemas asociados con la diversidad de APIs existentes. Por ejemplo, el estándar OpenSocial, recientemente trasladado al W3C (2014), incluye la descripción de distintas APIs abiertas que permiten a los desarrolladores acceder de forma unificada a los datos personales almacenados por distintos proveedores que implementan dicho estándar. Por otra parte, también existen soluciones propietarias o del ámbito de la investigación que tratan de unificar las llamadas a distintas redes sociales existentes como, por ejemplo, (Gouriten y Senellart, 2013).

A pesar de la diversidad de APIs existentes, especialmente en el contexto web, el estilo arquitectónico *Representational State Transfer* (REST) (Fielding, 2000) es ampliamente utilizado para permitir la comunicación entre APIs. Éste se basa en el *Hypertext Transfer Protocol* (HTTP) para dar conectividad y da libertad a los desarrolladores para especificar la sintaxis de los mensajes. Generalmente, se utiliza el formato de intercambio de datos *JavaScript Object Notation* (JSON) (Internet Engineering Task Force – IETF, 2014).

Las redes sociales son un ejemplo ilustrativo de lo expuesto anteriormente. Facebook y Twitter incluyen distintas APIs (Graph API¹⁴ y REST API¹⁵, respectivamente) que permiten leer y escribir datos personales de los usuarios, basándose en el protocolo HTTP y el formato de intercambio JSON utilizando. Aunque ambas soluciones se basan en las mismas tecnologías, dichos proveedores utilizan distinta sintaxis y semántica para referir a los mismos datos.

¹² Apple iTunes, <https://itunes.apple.com/us/genre/ios/id36?mt=8>

¹³ Google Play, <https://play.google.com/store>

¹⁴ Facebook Developers, The Graph API, <https://developers.facebook.com/docs/graph-api>

¹⁵ Twitter, REST API, <https://dev.twitter.com/rest/public>

3.3.4 Mecanismos de suscripción

Los mecanismos de suscripción permiten la recuperación de datos personales a medida que éstos se generan y/o actualizan en la fuente de datos. De esta forma, una entidad externa se suscribe a un campo personal del usuario y cuando dicho campo cambia, recibe las notificaciones o actualizaciones oportunas de la fuente de datos.

Generalmente, los mecanismos de suscripción están asociados al término **evento** que puede definirse como el cambio de estado dentro de un sistema o dominio particular que merece atención de otros sistemas o entidades (Taylor, et al., 2009). En el caso particular de datos personales, los eventos pueden incluir información sobre las acciones que realizan los usuarios, los datos que generan dentro de un determinado proveedor o los cambios de contexto o cualquier otra característica dinámica y temporal de los usuarios.

Además, el término evento puede referir a los hechos o sucesos que se producen (las instancias particulares de un tipo de evento) o a la comunicación de un evento específico (también denominado la notificación de una instancia de un evento).

Existen distintas tecnologías que permiten llevar a cabo la recuperación de datos personales a través de mecanismos de suscripción. Por ejemplo, los servicios de redes sociales incluyen distintas APIs como la Graph API de Facebook (en concreto, su funcionalidad llamada *Real Time Updates*¹⁶) o las Streaming APIs¹⁷ de Twitter que permiten recuperar sus recursos a medida que éstos se producen. En el primer ejemplo, la entidad solicitante configura los avisos que desea recibir sobre un determinado campo del perfil de usuario (por ejemplo, nombre o publicaciones creadas por el usuario) y cuando se produce un cambio en éste, Facebook lo notifica a través de métodos POST del protocolo HTTP. Las Streaming APIs de Twitter son ligeramente diferentes y requiere que se mantenga abierta una conexión HTTP de forma persistente.

Los mecanismos de suscripción están sustentados por el paradigma de Arquitectura Dirigida a Eventos (en inglés, *Event-Driven Architecture* – EDA) que se detallada a continuación.

3.3.4.1 Arquitectura dirigida a eventos

EDA es un patrón de arquitectura software que se caracteriza por establecer una comunicación asíncrona entre los actores participantes. A diferencia del modelo de interacción tradicional, como el que se constituye a través de los mecanismos de petición descritos anteriormente, donde una entidad solicita un recurso determinado y espera una respuesta, EDA se apoya en la generación de eventos que se transmiten sin una pregunta previa (aunque sí bajo unas condiciones preestablecidas de recuperación) y sin el conocimiento de lo que puede ocurrir con éstos posteriormente.

Principalmente, se asocia a este patrón ventajas relacionadas con el bajo acoplamiento entre los participantes, la facilidad para la reutilización de información, la automatización de

¹⁶ Facebook, Receiving API Updates in Real-Time, <https://developers.facebook.com/docs/graph-api/real-time-updates/v2.4>

¹⁷ Twitter, The Streamings APIs, <https://dev.twitter.com/streaming/overview>

procesos, la distribución de su arquitectura y el tratamiento y procesamiento de datos, entre otros (Malekzadeh, 2010).

La siguiente figura representa una arquitectura EDA genérica, constituida por tres niveles principales: productores, canal y consumidores.

Figura 3-7. Niveles de una arquitectura dirigida a eventos genérica (San Miguel, et al., 2014b).

Productores

Los productores representan la fuente de datos o eventos y así, son los encargados de detectar, crear y mandar eventos a través de un canal de comunicación. Éstos pueden contener **preprocesadores** para añadir inteligencia a la generación y realizar tareas como, por ejemplo, el filtrado, la priorización o la homogeneización de los eventos.

Existe una gran multitud de productores y con ello, de tecnologías asociadas a éstos. Todo ellos se enfrentan a uno de los retos más importantes de una arquitectura EDA: la definición de los eventos, lo cual incluye el formato a utilizar para las notificaciones y la semántica contenida en éstas.

No se ha detectado un estándar único o de facto para la definición y generación de eventos, ya que esto depende de la fuente de datos considerada. Por ejemplo, existen soluciones basadas en servicios web (como el *Web Services Notification*¹⁸ o el *Web Services Eventing*¹⁹), objetos Java, lenguajes de intercambio de datos (XML, JSON o *Google Protocol Buffers*²⁰), etc.

Canal

El **canal** lleva a cabo el transporte de las notificaciones de eventos entre los productores y los consumidores asociados. Normalmente, un canal toma la forma de Middleware Orientado a Mensajes (en inglés, *Message Oriented Middleware – MOM*) que envía y recibe mensajes entre sistemas, independientemente de las tecnologías, plataformas y recursos utilizados por éstos.

¹⁸ OASIS, Web Services Notification, https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsn

¹⁹ W3C, Web Services Eventing, <http://www.w3.org/Submission/WS-Eventing/>

²⁰ Google Developers, Protocol Buffers, <https://developers.google.com/protocol-buffers/?hl=es>

Un MOM puede aplicar diferentes modelos de mensajería como, por ejemplo, punto a punto o basados en publicación/suscripción. En el primer modelo, únicamente un consumidor recibe la notificación del evento, mientras que en el modelo de publicación/suscripción más de un consumidor puede expresar su interés en un tipo de evento para ser notificado (Eugster, et al., 2003). Técnicamente, lo anterior se alcanza gracias a una entidad intermediaria denominada como bróker de mensajería que recibe todas las notificaciones de eventos y las dirige a los consumidores suscriptos utilizando distintas colas de mensajería que almacenan las notificaciones si es necesario. Algunos ejemplos de estas soluciones son: RabbitMQ²¹ o ActiveMQ²², ambas basadas en el protocolo *Advanced Message Queue Protocol* (AMQP)²³.

Consumidores y motores de procesamiento de eventos

En último lugar, los **consumidores** son las entidades que reciben las notificaciones de eventos producidas y reaccionan ante éstas. Por ejemplo, cuando un consumidor recibe un evento puede crear nuevas notificaciones, invocar un servicio, iniciar un proceso de negocio, notificar a un humano para que realice una tarea manual, etc.

Existe un tipo especial de consumidor de eventos que dada su relevancia en ocasiones se representa como un nivel adicional de una arquitectura EDA: los **motores de procesamiento de eventos**. Éstos permiten realizar un conjunto de operaciones sobre los eventos recibidos como, por ejemplo, la lectura, creación, transformación, correlación, etc. (Etzion y Niblett, 2010).

Generalmente, se distinguen tres tipos de motores de procesamiento de eventos (Michelson, 2006):

- **Simple:** se produce cuando llega un evento y éste provoca el inicio de una acción. Normalmente, se utiliza para dirigir el flujo de eventos en tiempo real.
- **En *stream*** (o basado en corriente): llegan eventos más importantes o relevantes junto con otros más ordinarios o habituales y se analizan para ver el grado de importancia y en consecuencia, lanzar la ejecución de determinadas acciones. Generalmente, Se utiliza para tratar con grandes volúmenes de flujos de información en tiempo real y tomar decisiones sobre los mismos.
- **Complejo** (en inglés, *Complex Event Processing* – CEP): trata de evaluar la confluencia de los eventos que se reciben y en función de ello, lanzar la ejecución de acciones. Se lleva a cabo un análisis avanzado y una correlación entre eventos que puede ser causal, temporal o espacial. En este tipo de procesamiento es clave mantener el contexto y requiere la definición y marcado de patrones, técnicas sofisticadas de correlación, etc. Se utiliza normalmente para detectar anomalías de negocio, amenazas u oportunidades.

²¹ RabbitMQ, <https://www.rabbitmq.com/>

²² ActiveMQ, <http://activemq.apache.org/>

²³ AMQP, <https://www.amqp.org/>

Los motores de procesamiento de eventos pueden llevarse a cabo a través de distintos lenguajes como, por ejemplo, imperativos, basados en reglas de producción, reglas lógicas, etc. (Etzion y Niblett, 2010). Además, existen distintas herramientas comerciales y *open source*, como Drools Fusion²⁴ o EsperTech²⁵ que cuenta con su versión de código abierto Esper, que ofrecen procesadores de eventos simples, basados en stream o complejos.

3.3.5 Resumen

En esta sección se han revisado cuatro mecanismos principales que permiten llevar a cabo la recuperación de datos custodiados por proveedores de servicios: programas araña, ficheros estáticos de datos, basados en petición y basados en suscripción. La utilización de cada uno de ellos depende principalmente del tipo de acceso a los datos considerados (público o privado) y de los servicios específicos que el proveedor de dichos recursos habilite para la recuperación.

Cabe señalar que la recuperación de datos privados generalmente requiere que una entidad competente (el usuario al que refieren dichos datos o el propio proveedor) dé su autorización para llevar a cabo su recuperación, mientras que para los datos públicos puede llevarse a cabo sin ningún permiso especial.

Dentro de este orden de ideas, en el Capítulo 6 se presentan las contribuciones de la tesis doctoral respecto a la recuperación de datos. En concreto, se identificarán y caracterizarán distintas fuentes de datos, lo que permitirá establecer los componentes necesarios para la utilización y combinación de los mecanismos de recuperación descritos.

3.4 Gestión de identidad

La ITU (2009) define la **gestión de la identidad** como el conjunto de funciones y capacidades (por ejemplo, administración, gestión y mantenimiento, descubrimiento, intercambios de comunicación, correlación y vinculación, cumplimiento de una política, autenticación y asertos) que se utilizan para: garantizar la información de identidad (identificadores o atributos), garantizar la identidad de una entidad (por ejemplo, usuarios, dispositivos electrónicos u organizaciones) y apoyar aplicaciones de negocios y de seguridad.

Dada la definición anterior, los sistemas de gestión de la identidad abordan distintos retos relacionados con la creación, mantenimiento y uso de la identidad digital de los usuarios en cada una de las entidades participantes, por ejemplo, los proveedores de servicios. Esto incluye diversos servicios relacionados con la autenticación (certificación y verificación de la identidad digital), la autorización y control de acceso a recursos, la gestión del ciclo de vida de una identidad, la confianza entre las entidades participantes, la seguridad en los procedimientos realizados o la privacidad, entre otros, (Cao y Yang, 2010).

Simplificando todos los procedimientos y servicios que cubre un sistema de gestión de identidad, éstos engloban dos partes fundamentales: autenticación y autorización. De esta

²⁴ Drools, <http://www.drools.org/>

²⁵ EsperTech, <http://www.espertech.com/products/esper.php>

forma, permiten llevar a cabo la autenticación de una identidad, que tras una fase de registro, permite acceder a servicios y recursos ofrecidos, así como gestionar autorizaciones de acceso a terceras entidades. En esta tesis doctoral cobra especial importancia los mecanismos relacionados con las tecnologías para la delegación de autorización a través de la cual los usuarios permiten a terceras entidades actuar en su nombre, por ejemplo, para acceder a sus datos personales. Asimismo, la privacidad es de especial relevancia y ésta se tratará en una sección independiente (ver 3.5 Privacidad).

A continuación, se describen los modelos de gestión de identidad existentes y seguidamente, las tecnologías para la delegación de acceso relacionadas con la autorización a recursos personales de los usuarios.

3.4.1 Modelos de gestión de identidad

La gestión de la identidad puede abordarse a través de distintos modelos en los que participan tres entidades principales (ITU, 2009):

- **Usuario o sujeto:** solicita servicios de distintas entidades y para ello se apoya en la utilización de identidades digitales que permiten llevar a cabo las transacciones u otros propósitos.
- **Proveedor de identidad** (del inglés, *Identity Provider* – IdP): crea la identidad digital de un usuario y asegura la veracidad de la misma.
- **Proveedor de servicio** (o también denominados en inglés, *Relying Party*): ofrece al usuario servicios y para ello, se apoya en el uso de su identidad digital.

Dependiendo de las relaciones que se creen entre las entidades mencionadas, se pueden establecer distintos modelos que permiten llevar a cabo la gestión de la identidad. En este sentido, se pueden identificar cuatro modelos conceptuales principales (Jøsang y Pope, 2005) (Cao y Yang, 2010): aislado, federado (también denominado basado en red), centralizado y centrado en el usuario. A continuación, se describen éstos y por último, se incluye un resumen de las características halladas.

3.4.1.1 Modelo aislado

Tradicionalmente, los proveedores de servicios han llevado a cabo un **modelo de gestión de identidad aislado** en el que únicamente participan dos entidades: un proveedor de servicios, que también actúa como proveedor de identidad, y los usuarios. En este modelo (ver Figura 3-8), los usuarios establecen relaciones uno a uno con los proveedores con los que interactúa y en consecuencia, crean distintas identidades y credenciales (típicamente, usuario y contraseña) en cada uno de ellos.

Figura 3-8. Modelo de gestión de identidad aislado (Jøsang y Pope, 2005).

La principal ventaja del modelo de gestión de identidad aislado es que los proveedores de servicios pueden implementarlo de forma relativamente sencilla. Sin embargo, a medida que han ido apareciendo nuevos proveedores de servicios y aumentando el número de servicios telemáticos que utilizan los usuarios, éstos deben recordar múltiples identificadores y credenciales que pueden confundir u olvidar cuando quieren interactuar con un proveedor concreto. Además, para evitar lo anterior, los usuarios pueden llevar a cabo prácticas que comprometen la seguridad de su identidad como, por ejemplo, la elección de contraseñas débiles o fáciles de averiguar o la reutilización de contraseñas para distintos servicios. Todo lo anterior tiene un coste asociado para los proveedores de servicios, puesto que deben implementar mecanismos que permitan a los usuarios recuperar sus credenciales de acceso, establecer contraseñas seguras u otras acciones para garantizar la identidad de los usuarios.

En definitiva, conforme aumenta el número de servicios telemáticos que utiliza el usuario, el modelo de gestión aislado se convierte en poco usable para los usuarios y supone un esfuerzo añadido para los proveedores de servicios. Con el objeto de solventar lo anterior, han aparecido otros modelos de gestión de identidad como el federado, centralizado o centrado en el usuario, que se describen a continuación.

3.4.1.2 Modelo federado

El modelo de gestión de identidad federado consiste en un conjunto de acuerdos, estándares y tecnologías que permite a un grupo de proveedores de servicios reconocer las identidades digitales de los usuarios y los derechos de otros proveedores de servicios dentro de un dominio federado de confianza (Jøsang y Pope, 2005). Aquí, al igual que en el modelo aislado, cada proveedor de servicios mantiene las identidades de sus usuarios, sin embargo se establece una relación o mapeo con el resto de identidades de otros proveedores para que un usuario pueda ser reconocido en cualquiera de los proveedores de servicios federados. De esta forma, cuando un usuario se autentica en cualquier proveedor puede acceder al resto de forma segura y sin volverse a autenticar.

Desde el punto de vista del usuario, el modelo federado permite establecer una única identidad digital virtual y llevar a cabo un proceso de *Single Sign-On* (SSO) mediante el cual

únicamente necesitan autenticarse en una entidad y con esta autenticación podrán acceder directamente al resto de entidades federadas (ver Figura 3-9).

Figura 3-9. Modelo de gestión de identidad federado (Jøsang y Pope, 2005).

Las primeras especificaciones de sistemas de gestión de identidad federados fueron establecidas dentro del proyecto Liberty Alliance²⁶, actualmente sucedido por la iniciativa Kantara²⁷.

Para la Liberty Alliance, los proveedores de servicios están asociados a dominios autorizados, llamados **círculos de confianza**, que se apoyan en especificaciones tecnológicas del proyecto (una arquitectura y un conjunto de protocolos) y en una serie de acuerdos de operación y de negocio que definen las relaciones de confianza entre los proveedores de servicios. Cada proveedor de servicios mantiene los identificadores, contraseñas y atributos de los usuarios, sin embargo, los usuarios pueden asociar (o federar) dentro del círculo de confianza las cuentas aisladas que poseen en distintos proveedores de servicios.

Típicamente, un círculo de confianza está formado por un proveedor de identidad y distintos proveedores de servicios que confían en el primero. El proveedor de identidad permite autenticar a los usuarios, mantener las federaciones realizadas y emitir identificadores a otras entidades. Además, también pueden participar dentro del círculo de confianza proveedores de servicio de descubrimiento (en inglés, *Discovery Service* – DS) que saben dónde están almacenados los recursos de identidad del usuario y cómo acceder a ellos.

La Figura 3-10 representa un círculo de confianza donde se identifican las entidades participantes. Un ejemplo típico de aplicación de este modelo de gestión de identidad podría ser una compañía aérea que actúa como proveedor de identidad para distintas empresas asociadas a su programa de fidelización. Éstos son proveedores de servicios dentro del círculo de confianza definido.

²⁶ The Liberty Alliance, <http://www.projectliberty.org/>

²⁷ Iniciativa Kantara, <http://kantarainitiative.org/>

Figura 3-10. Gestión de identidades federada de las especificaciones del proyecto Liberty (Yelmo, et al., 2009).

Una de las piezas claves del modelo de gestión federado es el protocolo de federación que permite a distintos proveedores de servicios el intercambio de información de autorización y autenticación de forma segura. Por ejemplo, el *Security Assertion Markup Language* (SAML), un estándar abierto de la *Organization for the Advancement of Structured Information Standards* (OASIS, 2005), soportado por el proyecto Liberty. También el *Web Services Federation Language* (WS-Federation) (OASIS, 2009) asociado a la especificación de seguridad en servicios web (WS-Security) realizadas por IBM y Microsoft, entre otros.

A pesar de las ventajas que ofrece el modelo de gestión de identidad federado, éste tiene problemas similares a los señalados para los modelos aislados, puesto que los usuarios deben gestionar su identidad en distintos dominios de confianza y esto tiene asociado problemas de escalabilidad, ya que a medida que aumentan los dominios de confianza a los que pertenece el usuario, aumenta el número de identificadores y contraseñas que los usuarios deben recordar. Además, el modelo federado tiene asociados distintos retos para los proveedores de servicios relacionados con dificultades de integración, dadas las diferencias entre los mecanismos de seguridad, control de acceso o niveles de confianza (Almeshekah, 2010).

3.4.1.3 Modelo centralizado

A diferencia del modelo anterior, en el modelo de gestión de identidad centralizado, existe un único proveedor de identidad donde el usuario registra una identidad digital que es utilizada por el resto de proveedores de servicios (ver Figura 3-11). De esta forma, los proveedores de servicios no almacenan información de identidad de los usuarios y para llevar a cabo la autenticación o para acceder a la información del usuario recurren al proveedor de identidad establecido.

Figura 3-11. Modelo de gestión de identidad centralizado (Jøsang y Pope, 2005).

Existen distintos enfoques para poner en práctica el modelo centralizado como los indicados en (Jøsang y Pope, 2005). Sin embargo, cada vez más, están apareciendo proveedores de servicios (por ejemplo, Facebook, Google, LinkedIn²⁸ o Twitter, entre otros) que tratan de posicionarse como repositorios centrales de identidad en Internet aprovechando el gran volumen de usuarios que abarcan. Estas soluciones, denominadas como **social login** o **social sign-in** (Gafni y Nissim, 2014), se basan en protocolos abiertos como OpenID, y en la autorización delegada para el acceso a recursos de identidad a través de protocolos como OAuth (ver 3.4.2 Tecnologías para la delegación de autorización).

El estándar abierto OpenID (OpenID Foundation, 2007) define la secuencia de interacciones entre las distintas entidades participantes (ver Figura 3-12): usuario (en la figura representados como *User Browser*), proveedor de servicios (*Website A*) y proveedor de identidad (*OpenID Provider*). Dicha secuencia se basa en redirigir a los usuarios que quieren acceder a un proveedor de servicios y que soporta el protocolo OpenID al proveedor de identidad para completar el proceso de autenticación.

²⁸ LinkedIn, <https://www.linkedin.com/>

Figura 3-12. Procedimiento de autenticación en OpenID (Martin y Gil, 2012).

A pesar de que el protocolo OpenID tuvo una gran acogida inicial, éste ha recibido muchas críticas relacionadas con la complejidad de su implementación, la seguridad, privacidad y usabilidad y hoy en día, la OpenID Foundation²⁹, responsable de su promoción y administración, lo clasifica como una especificación obsoleta. En los últimos años, los mayores esfuerzos relacionados con el estándar han dado lugar al protocolo OpenID Connect (OpenID Foundation, 2014) que trata de solventar algunos de los problemas asociados. Las soluciones de los proveedores como, por ejemplo, Inicio de Sesión de Facebook³⁰, Google Sign-in³¹ o Sign in with Twitter³², basan sus soluciones en las nociones de los protocolos anteriores.

La siguiente figura muestra un ejemplo del proveedor de servicios Forbes³³ que permite el acceso de nuevos usuarios a través de las credenciales de Facebook, Twitter, LinkedIn, etc. (parte izquierda de Figura 3-13) u otras propias creadas por el usuario (parte derecha).

²⁹ OpenID Foundation, <http://openid.net/>

³⁰ Facebook, Inicio de Sesión, https://developers.facebook.com/docs/facebook-login?locale=es_ES

³¹ Google Sign-in, <https://developers.google.com/identity/>

³² Twitter, Sign in with Twitter, <https://dev.twitter.com/web/sign-in>

³³ Forbes, <http://www.forbes.com/>

Join Forbes using your social account:

Log in

Twitter

LinkedIn

Google

Yahoo

Microsoft

Social by Gigya

Sign up for a Forbes account below:

Name

This will be your display name on Forbes

Email

Choose a Password

Figura 3-13. Ejemplo de registro a través del *social login*.

Existen distintas compañías especializadas, como, por ejemplo, Janrain³⁴ o Gigya³⁵, que ofrecen soluciones para la implementación del modelo centralizado utilizando más de veinte proveedores de identidad diferentes. No obstante, tal y como refleja la Figura 3-14, Facebook es uno de los proveedores de identidad favoritos de los usuarios.

Figura 3-14. Proveedores de servicios elegidos por los usuarios para la autenticación en dispositivos móviles (parte izquierda) y la realización de compras electrónicas (parte derecha) en el segundo cuatrimestre del año 2015 (Gigya, 2015).

En definitiva, el modelo de gestión centralizado permite a los usuarios mantener una única identidad que puede ser utilizada por distintos proveedores de servicios y por consiguiente, mejora los problemas de usabilidad asociados al aumento de identidades digitales del usuario. Sin embargo, este modelo tiene asociados riesgos derivados del almacenamiento de dicha identidad digital en una única entidad como puede ser los relacionados con la seguridad o la privacidad. Es decir, la identidad de un usuario está custodiada por un único proveedor de servicios y de éste depende los mecanismos de protección y acceso a dicha identidad.

³⁴ Janrain, <http://janrain.com/>

³⁵ Gigya, <http://www.gigya.com/>

3.4.1.4 Modelo centrado en el usuario

El modelo de gestión de identidad centrado en el usuario (Bhargav-Spantzel, 2007) tiene como objeto ofrecer soluciones que dotan a los usuarios de capacidades para controlar y gestionar los procesos relacionados con la transferencia de su identidad digital, considerando una arquitectura que los sitúa en el centro y aspectos relacionados con la usabilidad que mejora la experiencia y facilidad de uso del sistema.

Desde que en 2007 se definiera por primera vez este modelo, han ido apareciendo distintos sistemas y soluciones que lo dan forma pero no está claramente establecido un enfoque único. Por ejemplo, en los primeros años aparecieron sistemas basados en tarjetas de identidad como los definidos por Higgins (Eclipse Foundation, 2009) y Windows CardSpace (Bertocci, et al., 2007). Éstos permiten a los usuarios gestionar sus identidades digitales bajo la metáfora de las tarjetas de visita del mundo real, apoyándose para ello en la utilización de un **agente de usuario** o una aplicación de confianza que se ejecuta en las máquinas y dispositivos de los usuarios. La siguiente figura muestra un ejemplo de este tipo de sistemas.

Figura 3-15. Ejemplo de un sistema de gestión de identidad centrado en el usuario (Jøsang y Pope, 2005).

Tomando la definición de la Information Card Foundation (2011), las tarjetas de identidad son la versión digital de las tarjetas que los usuarios pueden llevar en su billetera, que en el contexto digital se denomina como **selector**. Los selectores son el agente de usuario que permiten a los usuarios elegir las tarjetas que quieren utilizar frente a los distintos proveedores de servicios. Además, los atributos y características de éstas los define el usuario y el proveedor de identidad asegura que los datos introducidos son válidos, aunque la especificación no considera necesaria la participación de esta entidad en el modelo. De esta forma, los usuarios son los que deciden explícitamente qué datos se envían a qué proveedores, sin que haya un intercambio directo de recursos de identidad entre distintos proveedores de servicio o identidad como ocurren en las soluciones federadas.

La gran ventaja de este enfoque es que los usuarios tienen una experiencia de uso consistente, puesto que las interacciones siempre son las mismas a través del agente de usuario. No

obstante, en la práctica este enfoque no ha sido adoptado y lo más habitual es ver soluciones particulares ofrecidas por los proveedores de servicios o identidad (frecuentemente, junto con configuraciones relacionadas con el acceso a los recursos de los usuarios y la privacidad) que tienen características centradas en la supervisión del proceso por parte del usuario.

Por ejemplo, en la Figura 3-13, donde se observaba cómo el proveedor de servicios Forbes permite a sus usuarios autenticarse con distintos proveedores de identidad, puede considerarse una solución particular del modelo de gestión centrado en el usuario, puesto que se da libertad al usuario para decidir qué proveedor de identidad utilizar. Por otra parte, la Figura 3-16 muestra una captura de pantalla de una de las funcionalidades incluida por Facebook en el que un usuario puede ver y configurar qué aplicaciones o servicios acceden a qué datos y qué otros permisos tienen otorgados. En particular, se muestra un ejemplo para el proveedor de servicios Pinterest³⁶ aunque existe información de otros proveedores que se han apoyado en la utilización de la red social como proveedor de identidad.

Figura 3-16. Gestión de identidad en el proveedor de servicios Facebook.

3.4.2 Tecnologías para la delegación de autorización

Los modelos de gestión de identidad presentados anteriormente se han descrito resaltando las posibilidades que éstos ofrecen en cuanto a la identificación y autenticación de un usuario en un proveedor de servicios. Además, frecuentemente, estos sistemas incluyen capacidades relacionadas con la autorización a recursos personales que, tal y como define la ITU (2010), es una concesión de derechos y, sobre la base de esos derechos, concesión de acceso. Es decir, una autorización es un proceso independiente que garantiza o deniega el acceso a un recurso, al que generalmente le precede un proceso de autenticación.

³⁶ Pinterest, <https://es.pinterest.com/>

Existen distintas soluciones que permiten llevar a cabo la autorización a recursos personales de los usuarios que, generalmente, viene establecida por distintas condiciones y restricciones de acceso relacionadas con los recursos, el tiempo, lugar o contexto del acceso. En esta tesis doctoral, toman relevancia las soluciones que permiten a los usuarios delegar la autorización para el acceso a sus recursos personales a terceras entidades. A continuación, se describe el protocolo *Open Authorization* (OAuth) que ilustra la característica descrita.

3.4.2.1 Protocolo OAuth

El protocolo OAuth (IETF, 2012) especifica cómo una tercera entidad puede acceder a recursos personales de un usuario con su autorización. De esta forma, el propósito del protocolo es que un usuario que tiene determinados recursos personales alojados en un proveedor de servicios pueda dar acceso a un tercero a todos estos recursos (o únicamente, a parte de éstos) sin desvelar sus identificadores y credenciales, puesto que esto daría control total a su información de identidad.

A diferencia del sistema tradicional cliente-servidor basado en credenciales de acceso usuario-contraseña, el protocolo OAuth considera la participación de un nuevo rol denominado como **propietario de recursos** y el intercambio de claves de seguridad (denominadas como *tokens*). De esta forma, cuando un cliente (consumidor de recursos) desea acceder a un recurso del propietario (usuario), deberá obtener permiso del proveedor de recursos previa autorización del usuario. Dicha autorización se lleva a cabo en el proveedor del recurso y generalmente, está asociada a un proceso autenticación del usuario. En la siguiente figura se observa el flujo de interacciones concreto.

Figura 3-17. Visión global del protocolo OAuth (San Miguel, et al., 2015b).

El protocolo OAuth, o variantes de éste, es ampliamente utilizado como mecanismo de autorización por grandes proveedores de servicios como Facebook, Google o Twitter. Por ejemplo, el producto Inicio de Sesión de Facebook se basa en OAuth 2.0 y permite a los usuarios finales autorizar a distintos servicios (consumidores de recursos personales) el acceso

a los datos personales almacenados por la red social (proveedor de recursos personales) sin tener que dar sus credenciales a dichos consumidores.

3.4.3 Resumen

En esta sección se han descrito cuatro modelos conceptuales que permiten llevar a cabo la gestión de identidad: aislado, federado, centralizado y centrado en el usuario, así como las tecnologías para la delegación de autorización que permiten a los usuarios autorizar la recuperación de sus datos personales a terceras entidades. El establecimiento de éstos por parte de un proveedor de servicios responderá a sus capacidades y necesidades concretas, así como sus intereses de negocio.

Habitualmente, el modelo de gestión aislado aplica en contextos donde únicamente intervienen los usuarios y un proveedor de servicios que no requiere la interacción con otras entidades. Con el objeto de permitir la colaboración entre distintos proveedores y facilitar al usuario la gestión de sus identidades digitales en diferentes dominios administrativos, aparecieron los modelos de gestión federados, centralizados y centrados en el usuario. Principalmente, el primero se establece dado un acuerdo entre distintos proveedores, mientras que los modelos centralizados responden a un modelo de negocio en el que un proveedor se posiciona como entidad central para la gestión de la identidad frente a otros.

De forma transversal a los modelos de gestión mencionados, el modelo centrado en el usuario se basa en otorgar al usuario las capacidades para controlar y gestionar sus identidades digitales, bien entre distintos proveedores o en un mismo proveedor. Así, habilita los mecanismos y herramientas necesarias para que los usuarios decidan qué identidades o recursos personales se comparten con otras entidades y bajo qué circunstancias. Existen distintas tecnologías que permiten la realización de lo anterior como, por ejemplo, las tecnologías para delegación de autorización descritas en esta sección.

En el capítulo 8 se presentará una infraestructura para la gestión de la identidad que se basa en la combinación de tres modelos principales: centrado en el usuario, federado y centralizado. Además, dicha infraestructura permitirá establecer la gestión de la privacidad a los usuarios, un complemento esencial de la gestión de identidad.

3.5 Privacidad

A día de hoy, no existe una definición única consensuada para el concepto de privacidad, puesto que es un término complejo y plural que engloba distintos significados y percepciones relacionados con los derechos de las personas, la protección de datos, la intimidad y la gestión de identidad (Bertino y Takahashi, 2011) y envuelve aspectos culturales, sociales, legales, psicológicos, económicos y tecnológicos, entre otros.

Desde una perspectiva legal, la privacidad es un derecho fundamental de las personas recogido por la Declaración Universal de los Derechos Humanos (Naciones Unidas, 1948) en su artículo 12 y que dictamina: *“Nadie será objeto de injerencias arbitrarias en su vida privada, su familia,*

su domicilio o su correspondencia, ni de ataques a su honra o a su reputación. Toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques”.

En los últimos treinta años, el contexto en el que la privacidad debe ser entendida y aplicada ha cambiado sustancialmente (OECD ,2013). Los avances tecnológicos y en particular, Internet, han propiciado la aparición y generación de un gran volumen de datos personales que refieren directamente a las personas y que son recolectados, almacenados, usados y distribuidos día a día por múltiples actores del entorno digital. Por consiguiente, la privacidad ha ido extendiéndose y considerando nuevas situaciones y definiciones tal y como refleja Solove (2006). Tomando una de las más extendidas, la privacidad puede definirse como la acción de los individuos, grupos o instituciones para determinar cuándo, cómo y en qué medida sus datos personales se comunican a otros (Westin, 1968).

En definitiva, y tal y como indica (Abril y Pizarro, 2014), existen dos nociones prevalentes de la privacidad que deben considerarse en el contexto digital actual: **dignidad y control**. La primera refiere al respeto hacia la personalidad moral del individuo y sus derechos a mantener aspectos íntimos fuera del alcance de otros para proteger su integridad e independencia. Por su parte, la noción de control hace alusión a la libertad de los usuarios para decidir, bajo una decisión informada, quién accede a los datos personales que le refieren. Dentro de este orden de ideas, la denominada **integridad contextual** (Nissenbaum, 2004) ha aparecido como un marco sistemático para evaluar las expectativas de los usuarios respecto a la privacidad y las razones por las que ciertos hechos causan indignación moral, concluyendo que la privacidad tiene relación con las esferas y contextos en los que los datos personales se utilizan.

A lo largo del tiempo, numerosos gobiernos y organizaciones han propuesto distintos **principios** que establecen un conjunto de valores compartidos para la protección de la privacidad de las personas. Ejemplos de estos principios son: la guía de privacidad de la OECD (2013), los *Fair Information Practice Principles* (FIPs) de la Comisión Federal de Comercio (en inglés, *Federal Trade Commission – FTC*) (1998) de Estados Unidos o los incluidos en la Directiva 95/46/CE del Parlamento Europeo sobre la protección de datos personales. En términos generales, los ejemplos mencionados, aunque difieren desde una perspectiva legal, comparten las mismas ideas fundamentales. A modo ilustrativo, los principios básicos establecidos por la OECD (2013) son:

1. **Limitación de recogida.** Deberán existir límites en la recolección de datos personales y además, éstos deberán obtenerse por medios legales y en su caso, con el conocimiento o consentimiento de la persona a la que refieren.
2. **Calidad de los datos.** Los datos personales tendrán que ser correctos, completos, actuales y corresponder a los usos y fines para los que se van a utilizar.
3. **Especificación del propósito.** En el momento en que se produce una recogida de datos personales, deberá especificarse el fin que se va a dar a dichos datos y así, su uso posterior vendrá limitado por el cumplimiento de dicha especificación. Además, deberán especificarse en todo momento los cambios relacionados con el uso de los datos personales recolectados.

4. **Limitación de uso.** Los datos personales no se deberán usar para fines distintos a los especificados conforme al principio anterior excepto si se obtiene el consentimiento del usuario al que refieren dichos datos o por cumplimiento de un imperativo legal.
5. **Salvaguarda de la seguridad.** Los datos personales deberán estar protegidos contra riesgos relacionados con la seguridad, como, por ejemplo, el acceso no autorizado, la destrucción, modificación o revelación pública.
6. **Transparencia.** Se deberán establecer medios ágiles que permitan determinar la existencia, origen y usos de los datos personales, así como la identidad y lugar de residencia de la entidad que controla dichos datos con el objeto de ofrecer transparencia en el tratamiento, uso y otros aspectos relacionados con los datos personales.
7. **Participación individual.** Todas las personas tendrán derecho a conocer qué datos tiene una entidad sobre su persona en un tiempo razonable, a un precio no excesivo, y de forma sensata e inteligible. En el caso de que una petición sea denegada, las personas tendrán derecho a una explicación sobre las razones de dicha denegación y ésta se podrá cuestionar. De la misma forma, las personas podrán expresar sus dudas sobre el uso de sus datos personales y ejercer derechos relacionados con la eliminación, rectificación, modificación y corrección.
8. **Responsabilidad.** Las entidades que custodian los datos personales serán responsables directas del cumplimiento de las medidas llevadas a cabo para la aplicación de las directrices anteriormente descritas.

Los principios anteriores suponen la base o normas fundamentales sobre las que se debe regir la privacidad de un servicio, sin embargo no especifican cómo llevarlo a cabo. En este sentido, desde un punto de vista tecnológico, han aparecido multitud de mecanismos, tecnologías y herramientas cuyo objetivo es mejorar y fomentar aspectos particulares de la privacidad de las personas. Dichas herramientas se engloban bajo el término **Tecnologías para la Mejora de la Privacidad** (en inglés, *Privacy Enhancing Technologies* – PETs) (Goldberg, 2003) y dentro de éstas se incluyen soluciones dispares relacionadas con mecanismos de encriptación, protocolos de comunicación anónima, gestión de credenciales, utilización de pseudónimos, lenguajes de políticas de privacidad y su aplicación, etc. (Heurix, et al., 2015).

A pesar de las posibilidades y ventajas que las PETs ofrecen en el desarrollo de servicios telemáticos, éstas requieren un enfoque más substancial en el que se reconcilien y cubran las dimensiones legal y tecnológica de la privacidad (*European Union Agency for Network and Information Security* – ENISA, 2014). En este sentido, existe un consenso general por adoptar soluciones de diseño estratégicas que consideran la privacidad desde el comienzo del diseño de un servicio o producto, incluyendo el ciclo completo de su desarrollo. Dicho enfoque ha sido denominado como **Privacidad desde el diseño** (conocido por su término en inglés, *Privacy by design* – PbD) (Cavoukian, 2011) y se detallará en el apartado 3.5.1.

En los siguientes apartados se describen los conceptos relacionados con la privacidad desde el diseño y las distintas soluciones y tecnologías que permiten llevar a cabo la puesta en práctica de la privacidad, lo que se ha dado en llamar Ingeniería de la Privacidad (Spiekermann, 2009).

3.5.1 Privacidad desde el diseño

En los años 90, Ann Cavoukian, Comisionada de Información y Privacidad de Ontario, Canadá, acuñó el término privacidad desde el diseño, también traducido como privacidad por diseño, para atender los efectos crecientes y sistemáticos de los avances tecnológicos sobre las personas y sus derechos fundamentales. Progresivamente, el término ha ido cobrando una importancia y aceptación cada vez mayor y en la actualidad, sus principios han sido recogidos en distintos marcos legales como el proyecto de Reglamento General de Protección de Datos de la Unión Europea.

La privacidad desde el diseño rompe con los enfoques tradicionales para impulsar una situación en la que todos ganan y donde la funcionalidad y seguridad de los sistemas no está reñida con la protección de los datos personales (Privacy by Design, 2011). Para ello, la filosofía de esta doctrina es promover la introducción de la privacidad desde el inicio del diseño de un servicio o producto, considerando todas las áreas de aplicación de una organización, tecnologías de la información, prácticas de negocio, diseños físicos e infraestructuras de red (parte central en la Figura 3-18). De esta forma, la privacidad se convierte en el modo de operación predeterminado para las organizaciones.

Figura 3-18. Visión global del enfoque de privacidad desde el diseño (Privacy by Design, 2011).

La puesta en práctica de la privacidad desde el diseño gira en torno a **siete principios fundamentales** (ver Figura 3-18) (Cavoukian, 2011):

1. **Proactivo, no reactivo; preventivo no correctivo.** Este principio aboga por la anticipación y prevención de sucesos y eventos que supongan una invasión de la privacidad antes de que estos ocurran gracias a su consideración en el diseño. Así, tiene como objetivo evitar que se produzcan situaciones que vulneren la privacidad, más que esperar a que éstos se materialicen para resolverlos.

2. **Privacidad como configuración predeterminada.** La privacidad debe establecerse con niveles máximos y por defecto para asegurar automáticamente la protección de los usuarios. De esta forma, no se requiere que los usuarios lleven a cabo ninguna acción para configurar la protección de sus datos personales.
3. **Privacidad integrada en el diseño.** La privacidad debe entenderse como un componente esencial en el diseño de un sistema y es parte integral del mismo, de sus tecnologías, prácticas de negocio o infraestructuras físicas, sin disminuir su funcionalidad.
4. **Funcionalidad total, todos ganan.** Este principio defiende el equilibrio entre intereses contrapuestos, de forma que todas las partes involucradas ganan en el proceso. De este modo, no se renuncia a la protección de datos personales en pos de la funcionalidad o seguridad de estos recursos.
5. **Protección en el ciclo de vida completo.** La privacidad debe extenderse a lo largo de todo el ciclo de vida y no únicamente en el inicio, es decir, debe ser *extremo a extremo*, de principio a fin. Todos los datos personales deberán ser retenidos con seguridad y ser destruidos con la misma al final del proceso y sin demoras.
6. **Visibilidad y transparencia.** Todas las promesas y objetivos declarados de privacidad podrán estar sujetas, además de a la confianza, a la verificación por parte de los actores involucrados. Así, las acciones realizadas y establecidas en el sistema tienen que ser visibles y transparentes para los usuarios.
7. **Respeto y enfoque centrado en el usuario.** Los intereses de los usuarios deben estar por encima de todo, aplicando principios de usabilidad que faciliten la utilización de los servicios, informen y notifiquen de forma apropiada, etc.

Estos principios permiten establecer las características generales de los servicios y sistemas en cuanto a la privacidad, sin embargo, su puesta en práctica requiere la definición de métodos, técnicas y herramientas que permitan llevarlo a cabo de forma sistemática, lo cual ha sido denominado como Ingeniería de la Privacidad (Spiekermann, 2009), y se detalla a continuación.

3.5.2 Ingeniería de la privacidad

Actualmente no existe un enfoque tecnológico consensuado que permita trasladar los principios y conceptualizaciones relacionados con la privacidad en servicios telemáticos. De esta forma, a pesar de que existen múltiples herramientas PET que contribuyen a su mejora, la sistematización de la privacidad a través de soluciones tecnológicas se presenta como uno de los desafíos fundamentales en esta área de investigación.

La Ingeniería de la privacidad es una disciplina reciente cuyo objetivo es abordar los aspectos señalados y ofrecer soluciones relacionadas con la evaluación del impacto de la privacidad, el modelado de los flujos de datos, la gestión de los procesos asociados, el establecimiento de requisitos en el sistema, etc.

Tomando la taxonomía propuesta por (Spiekermann, 2009) existen dos enfoques principales complementarios para abordar la Ingeniería de privacidad: basado en políticas y basado en arquitectura. El primero se fundamenta en soluciones que consideran que los proveedores de servicios necesitan recolectar datos para llevar a cabo sus cometidos y por lo tanto, deben incluir sistemas que permitan notificar a los usuarios sobre sus conductas y conforme a esto, que los usuarios puedan decidir (*"notice and choice"* como lo denomina la autora del citado artículo). Por otra parte, las soluciones basadas en arquitectura se enfocan en mecanismos relacionados con la minimización en la recolección de datos personales, la anonimización de éstos y en general, con el almacenamiento y procesamiento de los datos.

En las siguientes secciones se presentan los trabajos más relevantes relacionados con el primer enfoque señalado que se ha abordado dados los requisitos del proyecto de investigación Segur@ donde se abordó esta temática.

3.5.2.1 Políticas de privacidad

Una política de privacidad es la definición de los requisitos, normas y reglas que permiten proteger el acceso a los datos personales de los usuarios y la divulgación de los mismos, así como los derechos particulares con respecto a cómo se utilizan dichos datos (ITU, 2010). Por consiguiente, ésta puede verse desde dos perspectivas diferentes principales: por una parte, permiten a las organizaciones expresar sus términos de uso y compartición respecto a los datos personales y por otra parte, facilitan a los usuarios la especificación de sus preferencias de privacidad.

(Madsen, et al., 2006) propone una taxonomía más detallada para las políticas de privacidad, donde además de considerar las preferencias de privacidad de los usuarios (como por ejemplo, *"Si yo doy mi dirección de correo electrónico, espero que..."*), tiene también en cuenta, desde el punto de vista de los proveedores de servicios, políticas que indican las promesas del proveedor (por ejemplo, *"Yo utilizaré el correo electrónico para..."*), las condiciones de aceptación (como *"No acepto el correo electrónico a menos que..."*), el gobierno de uso interno y compartición (por ejemplo, *"Sólo comparto el correo electrónico con terceros cuando..."*) y otras condiciones posteriores a la utilización o cesión de los recursos a terceros (como *"Si tú recibes el correo electrónico, debes destruirlo después de 3 días"*).

Las políticas de privacidad se expresan a través de distintos lenguajes, cada uno con su propia sintaxis y mecanismos de aplicación e implementación asociados. Ejemplos de éstos son: el *eXtensible Access Control Markup Language* (XACML), la *Platform for Privacy Preferences Project* (P3P), el *Client Attribute Requirement Markup Language* (CARML), *Enterprise Privacy Authorization Language* (EPAL), etc. (del Álamo, 2009), (Martínez, 2009).

A modo ilustrativo, a continuación, se presenta la infraestructura necesaria para la aplicación de políticas de privacidad conforme al estándar XACML, donde se identificarán los elementos necesarios, así como mecanismos asociados.

Una arquitectura para la aplicación de políticas de privacidad

XACML (OASIS, 2005b) es un estándar de OASIS que proporciona una arquitectura, un protocolo y un lenguaje completo basado en XML para la definición de distintas políticas de privacidad que rigen el intercambio de información entre proveedores de servicios.

Dentro de la arquitectura de XACML participan cuatro entidades principales, aunque dependiendo de los requisitos del sistema no todas tienen por qué estar presentes. Dichas entidades son:

- **Punto de Administración de Políticas** (*Policy Access Point – PAP*): entidad donde se crean y mantienen las políticas de privacidad. Éstas pueden estar almacenadas en un repositorio (*Policy Repository*).
- **Punto de Aplicación de Políticas** (*Policy Enforcement Point – PEP*): intercepta una solicitud de acceso y aplica la decisión del PDP.
- **Punto de Decisión de Políticas** (*Policy Decision Point – PDP*): evalúa y hace públicas las decisiones de autorización: permitido, denegado, indeterminado o no aplicable.
- **Punto de Información de Políticas** (*Policy Information Point – PIP*): consulta y recupera información solicitada por el PDP para que éste pueda tomar la decisión apropiada de autorización.

La Figura 3-19 muestra gráficamente la infraestructura que se establece en torno a las entidades descritas y la secuencia para la aplicación de políticas de privacidad. En primer lugar, se parte de distintas políticas establecidas en el PAP que se ponen a disposición del PDP. De esta forma, cuando una entidad (en la figura, *access requester*) realiza una solicitud de acceso que es detectada por el PEP, éste se dirige al PDP, que además puede apoyarse en el PIP, para emitir una respuesta a la solicitud. La comunicación entre el PEP, PDP y PIP se lleva a cabo a través del *context handler* que es el encargado de traducir las peticiones recibidas en el lenguaje XACML y viceversa, así como de obtener información de contexto necesaria para el PDP. El proceso finaliza cuando el PEP responde al solicitante con las obligaciones asociadas a la petición.

Figura 3-19. Arquitectura y secuencia para la aplicación de políticas de privacidad (Kim, 2009).

3.5.3 Resumen

Conforme a lo expuesto en esta sección, en los últimos tiempos, la privacidad se enmarca en un nuevo contexto en el que se requieren nuevas soluciones que permitan proteger y dar control a los usuarios sobre sus recursos personales originados en los servicios telemáticos.

En esta sección se han destacado distintos principios y disciplinas que tratan de conceptualizar la privacidad en el nuevo contexto como, por ejemplo, los establecidos por la OECD (2013) o el la Privacidad desde el diseño (Cavoukian, 2011). Además, se han descrito distintas soluciones que tratan de sistematizar los trabajos anteriores en el contexto tecnológico.

En el Capítulo 5 Especificación y diseño de un marco de datos personales de referencia se profundizará sobre la importancia de la privacidad en el contexto de esta tesis doctoral y su consideración desde el inicio del diseño. Además, posteriormente, en el Capítulo 8 Gestión de identidad y privacidad centrada en el usuario se describe una solución propuesta en esta tesis para la implantación de políticas de privacidad por parte del usuario.

3.6 Modelos de negocio en el ecosistema de datos personales actual

En esta sección se describen los modelos de negocio que han aparecido en torno a los datos personales de los usuarios o lo que es lo mismo, cuyo principal recurso, activo o valor del negocio deriva de éstos y que van más allá del propio procesamiento y aplicación que pueden llevar a cabo los proveedores de forma individual para ofrecer, por ejemplo, recomendación de contenidos, personalización en la interfaz o predicción del comportamiento de los usuarios. No obstante, en primer lugar, conviene definir y aclarar aspectos relacionados con el concepto de modelo de negocio y las posibilidades de éstos en el contexto de servicios telemáticos.

Finalmente, se describirán los modelos de negocio principales que se basan en el uso y gestión de datos personales.

3.6.1 Concepto y elementos del modelo de negocio

Dependiendo de dónde pongan el énfasis las distintas definiciones existentes, un **modelo de negocio** puede entenderse como: un sistema que permite la generación de valor para los clientes y la propia organización, un sistema de actividades interdependientes que conforman una arquitectura organizacional o la forma en la que una organización opera o refleja su estrategia realizada (López, 2012).

En el contexto de servicios telemáticos, y tomando una definición más formal, Timmers (1998) define un modelo de negocio como la arquitectura de un producto o servicio y sus flujos de información, incluyendo la descripción de los **actores** que participan en él y sus **roles**, el detalle de los beneficios potenciales para los diferentes actores participantes y la identificación de las fuentes de generación de ingresos. Adicionalmente, Chesbrough (2003) amplía la definición anterior e incluye los siguientes elementos como parte de un modelo de negocio:

- **Propuesta de valor:** permite identificar el valor creado para los usuarios a los que se les ofrece el producto propuesto y capturar valor para la organización que lo presta.
- **Segmentación de mercado:** incluye la identificación de los usuarios a los que va destinado el producto.
- **Cadena de valor:** permite crear y distribuir la oferta y los activos complementarios necesarios para soportar la posición del producto dentro de la cadena.
- **Mecanismos de generación de ingresos:** define cómo la solución propuesta obtendrá distintos pagos.
- **Estructura de costes y los beneficios potenciales:** se estima a partir de la propuesta de valor y la cadena de valor.
- **Red de valor:** identifica los competidores u otras empresas relacionadas con la solución propuesta y su vinculación con proveedores y clientes.
- **Estrategia competitiva:** permite establecer las ventajas con respecto a otros actores o competidores.

Existen otras metodologías, como la indicada en (Osterwalder y Pigneur, 2010), que considera, además de los elementos anteriores, otros elementos como los canales de comunicación, distribución y venta o las relaciones con los usuarios y clientes.

3.6.2 Modelos de negocio en el contexto de servicios telemáticos

Tradicionalmente, los modelos de negocio utilizados por los competidores dentro de un sector eran muy similares y el modelo de negocio en sí mismo no generaba grandes ventajas competitivas. Sin embargo, Internet y sus oportunidades asociadas como, por ejemplo, la

disminución de las barreras de entrada para la creación de un negocio, la reducción de costes de materias primas o asociadas con las ventas, la ampliación del alcance a clientes potenciales, la facilidad para llevar a cabo la identificación del segmento de mercado, etc., ha propiciado la aparición de nuevos modelos de negocio que han transformado las cadenas de valor tradicionales y reinventado los mecanismos de generación de ingresos en la mayoría de los sectores económicos.

Los servicios telemáticos pueden aplicar un modelo de negocio que no corresponde a única fórmula, sino a la combinación de distintos tipos de modelos y nuevas innovaciones. En la literatura, se pueden encontrar diferentes esfuerzos por catalogar los nuevos modelos de negocio surgidos, como puede apreciarse en (Zott, et al., 2010) que recopila distintas clasificaciones propuestas.

En esta tesis doctoral nos apoyaremos en la taxonomía propuesta por Rappa (2003), cuya clasificación se basa en la posición de un negocio dentro de la cadena de valor y los distintos mecanismos de generación de ingresos. Dada la rápida y constante evolución del sector, la taxonomía no pretende ser exhaustiva ni definitiva pero permitirá obtener una visión global de los principales modelos de negocios:

- **Modelo bróker o de corretaje.** Son negocios que crean valor de mercado, facilitando un espacio donde compradores y vendedores realizan distintas transacciones. Por lo general, obtienen ingresos por las transacciones realizadas pero las fórmulas son muy diversas. Dentro de esta clasificación se encuentran, por ejemplo, corredores de subastas (como eBay³⁷), plataformas de transacciones financieras (como PayPal³⁸) o mercados virtuales (como Amazon³⁹), entre otros. Los clientes a los que va dirigido el modelos pueden ser negocio a negocio, negocio a consumidor o consumidor a consumidor.
- **Modelo publicidad.** Es una extensión del modelo de negocio tradicional de los medios de comunicación. En este caso, el emisor es un sitio web que ofrece, normalmente de forma gratuita, contenidos o servicios (como blogs, correo electrónico, búsquedas, etc.) y donde se añaden mensajes publicitarios, como por ejemplo, *banners*, ventas emergentes o anuncios audiovisuales, que pueden ser el principal o único mecanismo de ingresos. Este modelo funciona mejor cuanto mayor es el volumen del tráfico de visitas o cuando la publicidad está más dirigida y especializada. Ejemplos de aplicación de este modelo son: portales de contenidos donde se incluyen anuncios (como Yahoo!⁴⁰), posicionamiento pagado que sitúa los anuncios en lugares destacables (como Google con su buscador), publicidad orientada al contenido (como Google en su servicio de correo electrónico⁴¹), anuncios con coste o pago por *click*, anuncios por cada mil impresiones, etc. Las posibilidades en este sentido son muy extensas.

³⁷ eBay, <http://www.ebay.es/>

³⁸ PayPal, <https://www.paypal.com>

³⁹ Amazon, <http://www.amazon.es/>

⁴⁰ Yahoo!, <https://es.yahoo.com/>

⁴¹ Gmail, <http://gmail.com/>

- **Modelo intermediario de la información (infomediarios).** El negocio ofrece mecanismos para que sus clientes comprendan la dinámica de un mercado determinado. Por ejemplo, servicios de medición de audiencia, marketing de incentivos (donde se crean programas de fidelización de clientes a través de distintos incentivos, como puntos o cupones de descuento, y los datos recogidos se ponen a la venta), servicios de generación de oportunidades de venta (basados en ofrecer bases de datos de clientes potenciales) y en general, cualquier negocio o servicio relacionado con la compra, venta y alquiler de datos de usuarios y especialmente, sus hábitos de consumo.
- **Modelo comerciante.** Este negocio se basa en mayoristas y minoristas de distintos bienes y servicios que ofrecen sus productos a través de precios fijos o por subastas. Por ejemplo, mercados virtuales (como Amazon³⁹), empresas que combinan sus establecimientos físicos con un sitio web de compras, catálogo virtuales, etc.
- **Modelo fabricante o modelo directo.** Permite crear un negocio directamente en Internet para ofrecer un producto o servicio y llegar así a los compradores, simplificando el canal de distribución. El éxito de su aplicación reside en la eficiencia y la comprensión de las preferencias de los clientes.
- **Modelo de afiliados.** Se basa en ofrecer incentivos financieros (generalmente, un porcentaje de ingresos) a los distintos negocios asociados o afiliados. Los ejemplos característicos de este modelos son los servicios de intercambio de *banners* (donde se negocia la colocación de éstos), el pago por *click* o reparto de ingresos.
- **Modelo comunidad.** Este modelo se establece gracias a la lealtad y fidelidad de los usuarios que se implican en un negocio, invirtiendo tiempo, recursos y emociones. Los ingresos económicos se basan en la venta de productos o servicios secundarios, contribuciones voluntarias o por la publicidad contextual. Existen muchos ejemplos de éxito de este modelo como los servicios de redes sociales, las comunidades de desarrollo de código libre o servicios de economía colaborativa, como por ejemplo, el micromecenazgo o *crowdfunding*.
- **Modelo suscripción.** En este modelo, los usuarios pagan periódicamente una cantidad por utilizar un servicio, aunque existen distintas variantes donde se ofrecen contenidos gratuitos y con contenidos de pago (*premium*) o se incluye publicidad. Este modelo se utiliza de forma masiva en proveedores de contenidos como, por ejemplo, la música o cine.
- **Modelo de utilidad.** Los negocios ofrecen un servicio o producto y facturan al usuario de acuerdo con la utilización real que éste realizada. Por ejemplo, servicios de telefonía como Skype⁴².

Actualmente, existe un conjunto de modelos de negocio que han resultado especialmente exitosos y cuya descripción conviene reflejar, puesto que pueden ser reutilizables y servir de

⁴² Skype, <http://www.skype.com/es/>

fuente de inspiración para la creación y establecimiento de nuevos modelos de negocio. En este sentido, en (Osterwalder y Pigneur, 2010) se identifican características y comportamientos similares que se han englobado en cinco **patrones de modelos de negocio** principales:

- **Desagregación de modelos de negocio.** Un modelo de negocio desagregado considera tres actividades empresariales independientes: relación con los clientes, innovación de productos e infraestructuras. Idealmente, cada una de estas actividades las asumen entidades independientes y esto da libertad para crear negocio en torno a nuevas actividades empresariales sin entrar en conflicto con el resto. La desagregación de modelos de negocio se ha comenzado a aplicar en el sector de las telecomunicaciones móviles en España, donde los operadores tradicionales comparten su recurso principal (la red) con competidores (operadores virtuales móviles) (por ejemplo, Movistar con los operadores virtuales Pepephone o Tuenti móvil⁴³) o subcontratan el mantenimiento de su infraestructura a fabricantes de equipos (como el acuerdo entre Yoigo y Ericsson⁴⁴). Además, algunos de estos operadores externalizan la innovación de productos en empresas más pequeñas y con ello, dinámicas, y esto les permite centrar su negocio en la relación con clientes (por ejemplo, Movistar y su programa de aceleración de *startups* Wayra⁴⁵).
- **La larga cola** (Anderson, 2006). El principio fundamental de este patrón es el de vender menos de más cosas, ofreciendo una gama de productos especializados que por separado tienen un volumen de ventas relativamente bajo pero cuyo conjunto puede resultar realmente rentable. El precursor de aplicación de este patrón es Amazon con su tienda virtual de todo tipo de productos.
- **Plataformas multilaterales.** En la actualidad, este patrón tiene una presencia y con ello, una importancia, cada vez mayor. Éste se basa en ofrecer una plataforma con dos o más propuestas de valor totalmente diferentes a distintos segmentos de clientes pero donde el valor del negocio no puede percibirse sin la existencia de alguno de los segmentos establecidos por las sinergias que se crean entre los mismos. Generalmente, el valor de estas plataformas aumenta con el aumento del número de clientes con el fenómeno que se conoce como “efecto de red”, y de igual forma, su éxito depende del número de clientes que haya en el resto de segmentos. Como se describirá posteriormente, Google es uno de los ejemplos más ilustrativos de este patrón de modelo de negocio, pero también ha sido aplicado en el sector de las consolas de videojuegos (donde se crean distintas sinergias entre los usuarios o jugadores y los desarrolladores de juegos) o el negocio de las tarjetas de crédito (que se dirige a los usuarios finales de éstas pero requiere de comercios que acepten dichas tarjetas para crear valor).

⁴³ Operadoras Móviles Virtuales (OMV), <http://operadoras-moviles.com/omvs/>

⁴⁴ Blog de Yoigo, noticia sobre la extensión del contrato de gestión de red, <http://yoigo.blogspot.com.es/2014/12/yoigo-y-ericsson-renuevan-contrato.html>

⁴⁵ Wayra, <http://wayra.co/>

- **Gratis como modelo de negocio.** Al igual que en el caso anterior, en este patrón generalmente participan distintos segmentos de clientes y al menos uno de ellos se beneficia constantemente de una oferta gratuita que, generalmente, financia el resto de los clientes participantes. Existen tres variantes principales: oferta gratuita basada en una plataforma multilateral (**publicidad**), servicios básicos gratuitos que se combinan con servicios *premium* opcionales (**modelos freemium**) y modelos donde se ofrece una oferta inicial gratuita o muy atractiva y que conllevan que determinados clientes realicen nuevas compras (**modelo del cebo y el anzuelo**). Los ejemplos más ilustrativos para los modelos mencionados son: Google y su modelo de negocio basado en publicidad, LinkedIn que se basa en modelos *freemium* ofreciendo servicios avanzados de pago y distintos juegos, como por ejemplo Candy Crush Saga⁴⁶, que permiten jugar gratuitamente pero cobran por obtener más vidas u otros productos relacionados con el avance de las partidas de juego.
- **Modelos de negocio abiertos** (Chesbrough, 2003). Este patrón permite crear y captar valor en un negocio con la colaboración de distintas entidades externas. Así, permite rentabilizar innovaciones internas y/o externas, bien ofreciendo recursos que no se están utilizando a agentes externos que pueden crear valor, o bien aprovechando recursos externos en los propios negocios. El paradigma de este patrón es en el que se sustenta la economía API mencionada en 3.3.3.1.

Además de los patrones descritos anteriormente, en los últimos años han cobrado fuerza distintas plataformas y proveedores de servicios que promueven la **economía colaborativa**. La Comisión Nacional de los Mercados y la Competencia (CNMC) (2014) define la economía colaborativa como un nuevo modelo de negocio que se basa en el intercambio entre particulares de bienes y servicios que permanecían ociosos o infrautilizados a cambio de una compensación pactada entre las partes. Por ejemplo, existen soluciones relacionadas con el transporte urbano como Urber⁴⁷ o interurbano como BlaBlaCar⁴⁸ que permiten poner en contacto conductores y pasajeros para compartir gastos, otras relacionadas con el alquiler de alojamiento turístico como Airbnb⁴⁹, donde particulares ofrecen sus casas, o servicios de financiación colectiva o en masa (conocidos por su términos en inglés, *crowdfunding*) como Kickstarter⁵⁰ que permiten financiar proyectos a través de pequeñas aportaciones individuales.

En términos generales, estos modelos de negocio eliminan intermediarios entre proveedores y consumidores, reduciendo costes en la transacción y se basan en la creación de conexiones sociales que fomentan la cooperación, localización, sostenibilidad y en definitiva, el empoderamiento individual y comunitario. Todo lo que está causando tensiones sobre los mercados en los que opera (CNMC, 2014).

⁴⁶ Candy Crush Saga, <http://candycrushsaga.com/>

⁴⁷ Urber, <https://www.uber.com/es/>

⁴⁸ BlaBlaCar, <https://www.blablacar.es/>

⁴⁹ Airbnb, <https://www.airbnb.es/>

⁵⁰ Kickstarter, <https://www.kickstarter.com/>

3.6.3 Modelos de negocio basados en datos personales

Los datos personales de los usuarios se han convertido en un recurso o activo económico fundamental en muchos de los modelos de negocio mencionados. El ejemplo más evidente son los modelos de negocio que se sitúan como intermediarios de información. No obstante, aunque no se determine explícitamente, el resto de los modelos descritos tienen sus mayores éxitos de aplicación cuanto más se adaptan y personalicen a las características particulares de los usuarios y sus clientes y con ello, cuantos más datos personales y conocimiento se posea de éstos. Lo anterior permite conectar con el usuario y/o cliente, mantenerlo, involucrarlo y en definitiva, fidelizarlo para garantizar el éxito de un negocio.

En los siguientes apartados, se describen los modelos de negocio más significativos respecto al uso y gestión de datos personales de los usuarios. Se ha obviado el modelo de negocio que pueden aplicar proveedores de servicios utilizando datos personales o conocimiento de los usuarios de forma independiente para mejorar las prestaciones del servicio o incluso, para prestarlo. Cabe señalar que no se entrará en la descripción de cada uno de los elementos de los modelos pero sí se destacarán sus características más relevantes y los distintos modelos y patrones aplicados, lo que permitirá obtener una visión global de cada uno de ellos.

3.6.3.1 *Brókeres de datos personales*

En la actualidad, existe una serie de compañías privadas, denominadas genéricamente como brókeres de datos (FTC, 2014), que aplican el **modelo de negocio de intermediario de la información** comercializando con datos personales de los usuarios que obtienen de distintas fuentes de datos públicas (como, por ejemplo, boletines del estado, blogs, perfiles de redes sociales, etc.) y/o comprando directamente a otras empresas que actúan como fuentes de datos privadas, ofreciendo, generalmente, información comercial de los usuarios como, por ejemplo, compras realizadas, cantidad, productos, etc.

Dado que el modelo de negocio de los brókeres de datos se basa en la recolección de datos personales de los usuarios de distintas fuentes de datos para generar valor, podría decirse que éstos aplican el patrón de modelos de negocio abiertos. Sin embargo, existe una falta de transparencia de las actividades realizadas por los brókeres de datos y por ello, se desconocen los detalles concretos de sus modelos de negocio. Un esfuerzo por aclarar esto se puede encontrar en (FTC, 2014).

En términos generales, el modelo de negocio de los brókeres de datos se basa en recolectar datos personales (información demográfica, historiales de navegación, datos de compras, datos financieros o de redes sociales, entre otros) que pueden procesar para derivar e inferir otros nuevos como, por ejemplo, las necesidades de compra específicas de una persona u otros más sensibles como la morosidad o la probabilidad de sufrir una enfermedad. Con todos los datos anteriores a su disposición, los brókeres los comparten o venden a otras compañías que los requieran (por ejemplo, otros brókeres de datos, instituciones financieras, compañías de seguro, hospitales, operadores de telecomunicaciones e incluso, entidades gubernamentales). Generalmente, los usuarios a los que se les atribuye los datos personales no participan en el modelo de negocio y desconocen su existencia.

La siguiente figura representa los actores participantes en el modelo de negocio, particularizado para distintas fuentes de datos de las que se obtienen datos personales y una única compañía a la que provee de datos el bróker (consumidor).

Figura 3-20. Modelo de negocio de los brókeres de datos personales.

Los brókeres de datos personales se han especializado en ofrecer productos dentro de tres ámbitos de aplicación principales: marketing (por ejemplo, para identificar clientes potenciales), mitigación de riesgos (para la detección de fraudes o la verificación de identidades) y búsqueda de personas (generalmente, un sitio web donde se puede consultar información agregada de otras personas que ha sido localizada, por ejemplo, en distintas redes sociales). En 2012, se cuantificó un volumen total de aproximadamente 426 millones de dólares en ingresos anuales, proviniendo la mayoría de los productos de marketing (192 millones) y la mitigación de riesgos (178 millones) (FTC, 2014).

3.6.3.2 Modelo de negocio de publicidad basado en el uso de datos personales

Hoy día, el **modelo de negocio de publicidad** es uno de los más consolidados en Internet y fundamentalmente, en aquellos servicios que ofrecen contenidos y/o servicios gratuitos a sus usuarios. Existen multitud de modelos de negocio basados en publicidad y cada uno de ellos corresponde a distintas composiciones con matices diferenciadores en los elementos del modelo de negocio. En este apartado nos centraremos en la descripción del modelo de negocio de publicidad genérico basado en el uso de datos personales, como puede ser el aplicado por grandes compañías como Google o Facebook.

El actor principal de este modelo de negocio (Figura 3-21) es un proveedor de servicios que aplica un patrón de **plataforma multilateral**, donde el proveedor se dirige a dos segmentos del mercado principales: los usuarios de sus servicios y los anunciantes o terceros proveedores de servicios que desean publicitarse. Por una parte, ofrece una propuesta de valor gratuita a sus usuarios, permitiendo el acceso a, por ejemplo, servicios de búsqueda de información, comunicación con otras personas, consumo y/o compartición de contenidos multimedia, consulta de mapas, etc. Por otra parte, el proveedor de servicios ofrece a los anunciantes publicidad dirigida de manera muy efectiva, utilizando, por ejemplo, características

demográficas, geolocalización o contenidos consultados por los usuarios. El éxito de este modelo se fundamenta en conseguir un gran número de usuarios que generen datos personales constantemente y así, conocer mejor a éstos.

Figura 3-21. Modelo de negocio de publicidad.

Generalmente, los ingresos económicos del proveedor de servicios provienen de los anunciantes y los mecanismos de pago pueden ser muy diversos (por ejemplo, pago por *click*, pago por número de impresiones, afiliados, suscripciones, etc.), aunque se puede combinar con otros mecanismos de ingresos (servicios *premium*, monedas virtuales, juegos, etc.).

3.6.3.3 Plataformas de datos abiertos

Las plataformas de datos abiertos ofrecen de forma gratuita datos de distinta índole, como información demográfica, meteorológica, genomas, tráfico, etc., para que terceras entidades (usuarios, empresas u otros organismos) puedan reutilizarla. El patrón de estas plataformas corresponde a los **modelos de negocio abiertos** y lo pueden aplicar tanto empresas privadas como públicas, sin embargo, uno de los ejemplos más ilustrativos son los que llevan a cabo distintas instituciones gubernamentales y organismos públicos.

Este modelo de negocio (Figura 3-22) consiste en exponer públicamente distintos datos disponibles en una organización de forma que se permita su acceso y libre reutilización a terceras entidades (reutilizadores). Dichas entidades pueden consumir estos datos y con ello, generar nuevos productos o servicios y con ello, valor. Salvo excepciones, el retorno económico obtenido por las plataformas de datos abiertos es difícil de cuantificar. A modo de ejemplo, el Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información (ONTSI) ha emitido distintos informes que tratan de analizar y cuantificar el sector, que denominan **sector infomediario**, y que engloba a las distintas empresas que se basan en la reutilización de información tanto del sector público y/o privado para la generación de nuevos productos o servicios. En (ONTSI, 2015) se señala que los datos provenientes del sector público generan un volumen de negocio de entre 450 y 500 millones de euros y además, profundiza en aspectos relacionados con los datos de origen del sector privado señalando las características de la información y las actividades reutilizadoras de la misma.

A pesar de lo difícil que es cuantificar el volumen generado por este modelo de negocio, lo que sí está demostrado es que las plataformas de datos abiertos producen distintos beneficios colaterales en el resto de actores como, por ejemplo, el crecimiento económico, la creación de nuevas empresas y modelos de negocio que explotan los recursos, el fomento de la innovación, el aumento de la confianza de los usuarios (por ejemplo, ciudadanos) al proveedor

de plataforma (en el ejemplo, la administración), la mejora de colaboración entre los actores del modelo, el aumento de eficacia y eficiencia (Álvarez, 2014). Todo lo anterior fomenta la creación de nuevos ecosistemas alrededor de las plataformas de datos abiertos.

Figura 3-22. Modelo de negocio de plataformas abiertas de datos.

En general, los datos que pueden ofrecer las plataformas de datos abiertos pueden ser de cualquier ámbito y formato de representación, sin embargo, en el caso de España, la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público especifica que deben respetar el derecho a la protección de datos personales y la propiedad intelectual. Así, en el caso concreto de los datos personales, éstos deben anonimizarse previamente.

3.6.3.4 Nuevos modelos de negocio basados en la participación del usuario

En los últimos años han comenzado a proliferar un conjunto de iniciativas y nuevas empresas que proponen nuevos modelos de negocio en torno a los datos personales. Los enfoques de estas soluciones son muy dispares y dado lo reciente de las mismas, no es posible profundizar en los detalles de los modelos de negocio de cada una de ellas. Sin embargo, se ha detectado una tendencia que se basa en romper la cadena de valor de tradicional de los modelos de negocio explicados anteriormente para involucrar al usuario. A continuación, se describen los aspectos más relevantes hallados.

En 2014 apareció Datacoup⁵¹, una de las primeras empresas en ofrecer directamente dinero a los usuarios a cambio de sus datos personales de distintas redes sociales y tarjetas de débito para vendérselos a terceras entidades. El servicio está disponible únicamente en Estados Unidos y ofrece una cantidad aproximada de 6-8 dólares al mes, aunque esto varía en función de los datos. Datacoup declara que los datos se anonimizan y agregan para vendérselos a compradores interesados. Por su parte, los usuarios pueden elegir qué datos se comparten y con quién.

⁵¹ Datacoup, <https://datacoup.com/>

De forma similar, distintos medios, por ejemplo (Ross, 2014), señalan que distintas compañías como Luth Research⁵² y Verizon⁵³ pagan o recompensan a los usuarios por monitorizar su comportamiento a través de sus dispositivos electrónicos. En el primer caso, se ofrecían hasta 100 dólares al mes por los datos del uso del portátil, ordenador o tablet, requiriendo la monitorización completa de dichos dispositivos.

Otras compañías no pagan directamente a los usuarios por aportar sus datos personales pero ofrecen distintas propuestas de valor que incentivan al usuario a aportar sus datos personales. Por ejemplo, Reputation⁵⁴ pone en contacto a empresas y usuarios para que estos últimos a cambio de determinados datos personales obtengan distintos descuentos.

Finalmente, otras iniciativas fomentan la participación del usuario promoviendo el bienestar social o el de determinados usuarios. Este es el caso del sitio TheGoodData⁵⁵ que obtiene los datos de navegación del usuario con su consentimiento para vendérselos a terceras entidades y los beneficios generados se reparten el 50% en proyectos de países subdesarrollados y el otro 50% al mantenimiento del servicio o la plataforma PatientsLikeMe⁵⁶ que se basa en recolectar directamente de los usuarios datos médicos sobre distintas enfermedades (tratamientos, síntomas, etc.) para facilitar la investigación a profesionales del sector y ayudar a otros pacientes.

3.6.4 Resumen

En esta sección se ha presentado un estudio sobre los distintos modelos de negocio que se han establecido en el contexto de los servicios telemáticos y en particular, sobre los que se han originado en torno a la utilización y compartición de datos personales de los usuarios. En el siguiente capítulo se presenta un análisis crítico de los modelos de descritos y sobre la base de éste se describe un modelo de referencia propuesto en esta tesis doctoral. Dicho modelo aspira a cubrir las debilidades y amenazas detectadas en el ecosistema de datos personales actual.

⁵² Luth Research, <https://www.luthresearch.com/>

⁵³ Verizon, <http://www.verizonwireless.com/>

⁵⁴ Reputation, <http://www.reputation.com/>

⁵⁵ TheGoodData, <https://thegooddata.org/>

⁵⁶ PatientsLikeMe, www.patientslikeme.com

Capítulo 4.

Nuevos modelos de negocio para el ecosistema de datos personales

En el Capítulo 1 se describió el ecosistema de datos personales actual de donde se desprende que cada uno de los participantes (usuarios, proveedores de servicios y el sector público) está expuesto a una serie de oportunidades y amenazas que resultan en ocasiones en situaciones paradójicas.

Por una parte, el sector público se enfrenta ante la difícil disyuntiva de promover iniciativas que permitan la generación de nuevo valor y a la vez, proteger a los usuarios o ciudadanos con legislaciones que velen por el uso de sus datos personales y en definitiva, su privacidad. Parece que en la economía actual, basada en el uso de estos datos y la generación y el aprovechamiento del conocimiento de los usuarios, reconciliar ambas posiciones supone abordar una tarea ardua. Por otra parte, los usuarios disfrutan de multitud de servicios telemáticos y no están dispuestos a renunciar a éstos. Los servicios telemáticos ofrecen grandes ventajas y avances en la sociedad, sin embargo, un uso inadecuado o inapropiado de los datos personales que los usuarios generan en dichos servicios se traduce en un riesgo para la privacidad del usuario. En este sentido, aparentemente existe una preocupación creciente de los usuarios respecto al uso de sus datos personales. Finalmente, en el ecosistema de datos personales actual se hallan grandes diferencias entre los proveedores de servicios, dependiendo principalmente de su volumen de negocio. Así, existen proveedores de servicios que se han implantado como casos exitosos gracias al uso, gestión y aprovechamiento de los datos personales de sus usuarios y otros proveedores que no cuentan con dichos recursos de antemano y cuya entrada en el sector puede resultar compleja y dificultosa.

Todo lo anterior nos permite concluir que el ecosistema de datos personales está fragmentado y es ineficiente (WEF, 2011) y por lo tanto, que para cada uno de los participantes existe una serie de amenazas y debilidades que desequilibran el ecosistema actual y que requieren soluciones específicas que permitan su evolución y reconciliación entre los mismos.

En este capítulo se establece el posicionamiento de la tesis doctoral desde una perspectiva económica, presentando un nuevo modelo de negocio de referencia que ofrece a los participantes más débiles del ecosistema de datos personales actual las herramientas y servicios necesarios para reforzar su presencia.

De esta forma, en primer lugar, en la sección 4.1 se incluye un análisis de los distintos modelos de negocio que actualmente se basan en el uso y la compartición de datos personales entre distintas proveedores de servicios (ver 3.6.3 Modelos de negocio basados en datos personales) y que permitirá detectar las fortalezas y debilidades de los mismos, así como dar una visión global del nuevo modelo de negocio propuesto. Seguidamente, en la sección 4.2 se describen los elementos del modelo de negocio propuesto y en la sección 4.3 se incluye la particularización del mismo en dos ámbitos de aplicación y en concreto, en el sector de publicidad en redes sociales y el sector financiero. Por último, la sección 4.4 incluye un resumen y las contribuciones originales incluidas en este capítulo.

4.1 Posicionamiento

En el Capítulo 3 Estado del arte, apartado 3.6.3, se revisaron los distintos modelos de negocio que han aparecido en torno a los datos personales y que van más allá de la utilización y aplicación individual de los datos personales de los usuarios por parte de los proveedores de servicios, ya que éstos se basan en el intercambio de datos personales con terceras entidades. Del estudio y análisis realizado, obtenemos las siguientes conclusiones:

- El modelo de intermediario de la información que aplican distintos **brókeres de datos** genera una serie de tensiones respecto a la ética, conductas y responsabilidades de su modelo de negocio (Toca, 2014). La falta de transparencia actual de estos proveedores dificulta conocer qué datos almacenan, a quién refieren, qué indican y mucho menos, dan a los usuarios la posibilidad de borrar o modificar los datos que poseen (muchas veces incorrectos o extremadamente sensibles como, por ejemplo, tendencias sexuales, posiciones políticas o posibilidades de sufrir una enfermedad) o conocer a quién o quiénes se vende dicha información. En este sentido, los brókeres de datos se están encontrando con una gran oposición y desconfianza desde Europa e incluso, Estados Unidos (FTC, 2014). En el caso particular de España, las actuaciones de los brókeres de datos personales podría considerarse ilegal atendiendo a la Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal.
- El **modelo de negocio de publicidad** basado en el uso de datos personales de los usuarios está en manos de muy pocas, pero muy potentes, compañías privadas, lo cual genera preocupaciones en torno al sector. El ejemplo más ilustrativo de dichas tensiones es el caso particular de la compañía Google con una situación de dominio y casi monopolística que resulta compleja de equilibrar dadas las sinergias que se han producido entre los distintos actores de su modelo. En Wikipedia (2015) se recopilan las noticias sobre las mayores críticas que se hacen a la compañía relacionadas con la utilización de cookies y el rastreo de usuarios, la captura de imágenes de su servicio de

mapas *Street View*⁵⁷, el derecho al olvido o la protección de menores. En definitiva, cada vez más, los proveedores dominantes de este modelo ofrecen a sus usuarios una oferta gratuita más atractiva y variada, lo cual aumenta el tráfico de visitas, usuarios, y con ello, la de anunciantes interesados en llegar a ellos a través de este canal. Todo ello genera barreras de entrada a nuevos participantes que difícilmente pueden superar los valores ofrecidos por los grandes proveedores de servicios.

- Las **plataformas de datos abiertos** suponen una gran oportunidad para promover la reutilización de datos de distinta índole y favorece la entrada de nuevos actores que pueden utilizar esta información para ofrecer valor añadido. Sin embargo, existe una falta de homogeneidad de los datos ofrecidos y dificultades de tratamiento y localización de éstos, lo cual puede actuar como barrera de entrada para nuevos actores (ONTSI, 2015) y (Networked & Electronic Media, 2013).
- En los últimos años han proliferado **nuevos modelos de negocio basados en la participación del usuario** para, principalmente, aportar datos personales a cambio de obtener algún beneficio. Las actividades y servicios de los proveedores de servicios que adoptan este modelo son muy variadas y se pueden encontrar distintas propuestas de valor que incentivan a los usuarios a participar. Sin embargo, no parece que esté claro si los usuarios realmente saben para qué se utilizan los datos que ellos mismos aportan.

Ante este contexto y dadas las consideraciones mencionadas, el posicionamiento de esta tesis doctoral se fundamenta en la introducción de un nuevo actor en el ecosistema de datos personales actual que permita cubrir las debilidades y amenazas señaladas. De esta forma, el modelo de negocio propuesto aprovecha las ventajas de los modelos de negocio actuales y promueve el avance de éstos para equilibrar los conflictos y perjuicios detectados.

En concreto, el actor principal del modelo de negocio es un nuevo proveedor de servicios o entidad que ha sido denominada como **marco de datos personales** (*Personal Data Framework – PeDF*) que aplica un enfoque de **plataforma multilateral** para dirigirse tanto a proveedores de servicios finales, como a los usuarios. Además, se apoya en un **modelo de negocio abierto** a través del cual se beneficia de los recursos que terceras entidades ofrecen de manera pública y/o privada y a su vez, provee a otras entidades de nuevo conocimiento de los usuarios con el que pueden a su vez producir nuevo valor y que, difícilmente, podrían alcanzar de forma unilateral. En este contexto, los usuarios tienen las herramientas necesarias que les permiten controlar el intercambio de sus datos personales y además, pueden recibir distintos incentivos por participar en el modelo de negocio.

En la siguiente figura se ha representado una visión global del modelo de negocio del marco de datos personales en relación a las representaciones de los modelos de negocio descritos en 3.6.3. Aquí, se puede observar cómo, al igual que lo brókeres de datos personales, el marco de datos personales recolecta recursos personales de distintas fuentes de datos, sin embargo, a diferencia de éstos, la recolección está supervisada por el usuario y se tiene su consentimiento. Además, el marco de datos personales provee a terceras entidades con

⁵⁷ Google Street View, <http://www.google.es/intl/es/maps/streetview/>

distintos recursos personales. Por una parte, de manera semejante a las plataformas de datos abiertos, puede ofrecer datos personales a distintos proveedores de servicios (reutilizadores) y lo que es más interesante y le diferencia de la mayoría de los modelos de negocio revisados, puede ofrecer nuevo conocimiento de los usuarios (modelos de usuario) a otras entidades (consumidores), todo lo anterior para generar nuevo valor que puede repercutir de nuevo en el marco de datos personales. Es importante destacar que la participación del usuario está presente en todas las actividades que sustentan el modelo de negocio, desde la recolección hasta la provisión de nuevos recursos personales y por ello, se verá beneficiado o recompensado.

Figura 4-1. Visión global del modelo de negocio del marco de datos personales.

En definitiva, el marco de datos personales se basa en el modelo de negocio de intermediario de la información para dirigirse tanto a proveedores de servicios, como a los usuarios. Éste puede realizarse en distintos ámbitos de aplicación dependiendo de los objetivos específicos de su proveedor y en este sentido, se han detectado tres contextos principales:

- **Sector público.** Desde el sector público, distintas organizaciones o instituciones pueden promover la realización de un marco de datos personales para proveer a los ciudadanos de las herramientas necesarias que permitan controlar el flujo de su información personal, a la vez que promueven y facilitan la generación de valor a otras entidades públicas o privadas y la entrada de nuevos actores que puedan aprovechar los recursos ofrecidos.
- **Sector privado.** Una empresa puede especializarse como proveedor de marco de datos personales en un determinado ámbito de aplicación como, por ejemplo, datos comerciales, comunicaciones móviles, banca, etc. y obtener beneficios económicos con el modelo de negocio propuesto.

- **Consortio de organismos y empresas.** Se pueden establecer distintos acuerdos entre entidades del sector público y privado para llevar a cabo un marco de datos personales y obtener un beneficio común. En este contexto, el proveedor del marco de datos personales sería la dirección administrativa del consorcio que coordinaría al resto de entidades.

En la siguiente sección se detallan los elementos del modelo de negocio sin tener en cuenta ningún ámbito de aplicación específico y siguiendo un enfoque generalista. Seguidamente, en la sección 4.3 se detallan dos ámbitos de aplicación que se han validado en distintos proyectos de investigación.

4.2 Modelo de negocio de referencia para un ecosistema de intercambio de datos personales centrado en el usuario

La entrada de la nueva entidad denominada en el ecosistema de datos personales actual da lugar a un nuevo modelo de negocio que sustenta la innovación presentada en esta tesis doctoral. El objetivo de esta sección es describir en detalle los distintos elementos del modelo de negocio propuesto para el marco de datos personales.

Como se señaló en 3.6 Modelos de negocio en el ecosistema de datos personales actual, los modelos de negocio están en constante evolución y no existe una fórmula mágica que garantice el éxito de un negocio. Los modelos de negocio deben ser revisados para que resulten pertinentes y eficaces en nuevos contextos y así, deben aparecer nuevas combinaciones e innovaciones que creen oportunidades y beneficios. En este sentido, el modelo de negocio del marco de datos personales supone una combinación de distintos modelos de negocio mencionados pero se pueden añadir variaciones y nuevos matices que permitan la aplicación en distintos ámbitos y su evolución en un futuro.

Partiendo de las definiciones del Capítulo 3 sección 3.6, en los siguientes apartados se describen los elementos principales del modelo de negocio que se propone para el marco de datos personales. En primer lugar, se identifican los actores participantes, los roles que estos asumen y su relaciones, seguidamente se describen las propuestas de valor que el marco de datos personales ofrece a sus clientes, la segmentación del mercado, la cadena y red de valor, y finalmente, los mecanismos de generación de ingresos, lo cual incluye los mecanismos de incentivos del modelo de negocio.

4.2.1 Actores, roles y relaciones

El modelo de negocio propuesto hereda las dos entidades o **actores principales** del ecosistema de datos personales actual:

- **Proveedores de servicios finales:** entidad del ámbito privado o público que ofrece a los usuarios servicios telemáticos de cualquier tipo como por ejemplo, compras electrónicas, comunicación en redes sociales, servicios financieros, consultas de datos

universitarios, etc. Sólo por proporcionar el servicio, y mucho más para ofrecer una experiencia de uso adaptada a los usuarios, requieren y tratan con datos personales de los usuarios. Además, como respuesta a la estrategia propia de su modelo de negocio (por ejemplo, para dar valor añadido a los usuarios, mejorar su imagen de marca o para que terceras entidades generen nuevo valor) los proveedores de servicios pueden incorporar los mecanismos necesarios para exponer sus recursos a terceras entidades. En particular, entre dichos recursos se pueden encontrar datos personales o conocimiento avanzado de los usuarios en forma de modelos de usuario.

- **Usuarios:** persona física que disfruta de los servicios ofrecidos por los proveedores de servicios finales y como consecuencia, aportan y generan datos personales en dichas entidades y, a su vez, éstas pueden derivar o inferir distintos modelos de usuario. En el modelo de negocio presentado, los usuarios cuentan con un proveedor de servicios más que les permite gestionar el flujo de sus recursos personales y así, establecer su control en el uso y la cesión de éstos.

Además de los anteriores actores, hay que añadir uno nuevo en torno al cual gira el modelo de negocio: el **proveedor del marco de datos personales**. Éste ofrece distintos servicios específicos al resto de actores del modelo, actuando como una entidad intermediaria entre proveedores de servicios finales y los usuarios.

Cada uno de los tres actores principales puede asumir distintos **roles** dentro del modelo de negocio, adquiriendo nuevas habilidades que permiten el uso y la gestión de distintos recursos personales (generalmente, datos personales y/o modelos de usuario), tal y como se ha planteado en la sección 4.1 Posicionamiento. En particular, los roles identificados son:

- **Proveedor de recursos personales:** expone y proporciona datos personales y/o modelos de usuarios a terceras entidades para que éstas puedan recuperarlos y llevar a cabo distintas actividades individuales. Este rol lo pueden adoptar todos los actores del modelo de negocio. En primer lugar, los proveedores de servicios finales pueden exponer los datos personales que poseen de sus usuarios y, aunque generalmente no es tan común, los modelos de usuario que han generado individualmente. También el proveedor del marco de datos personales puede ofrecer los datos recuperados de otros proveedores a terceras entidades pero lo más característico de éste es que ofrece modelos de usuario generados a partir de los datos personales recuperados de distintos proveedores de recursos. Finalmente, el usuario, con las herramientas y servicios adecuados pueden aportar datos personales e incluso, modelos de usuario sobre sí mismos.
- **Consumidor de recursos personales:** obtiene datos personales y/o modelos de usuario de los distintos proveedores de recursos para generar valor. Por ejemplo, el propietario del marco de datos personales asume este rol cuando consume datos personales de distintos proveedores de recursos personales para generar nuevos modelos de usuario que ofrece a terceras entidades. A su vez, los proveedores de servicios finales pueden requerir modelos de usuario del marco de datos personales y así, asumir el rol de consumidor de recursos personales. También los usuarios pueden

actuar como consumidores de recursos personales cuando obtienen, por ejemplo, datos personales de un proveedor de servicios para importarlo en otro proveedor de servicios como el marco de datos personales. Una especialización de este rol son los **reutilizadores** de recursos personales, que se basan en la generación de nuevos recursos personales a partir de los obtenidos de otras fuentes.

- **Generador de nuevos recursos personales:** obtiene, a partir de recursos personales propios o de aquellos recuperados en distintos proveedores, nuevos recursos personales (generalmente, modelos de usuario a partir de distintos datos personales). Los nuevos recursos generados se pueden utilizar para llevar a cabo distintas actividades individuales o para ofrecerlos a terceras entidades. Típicamente, este rol lo asume el proveedor del marco de datos personales para la generación de nuevo conocimiento de los usuarios que, posteriormente, ofrece a terceras entidades, aunque también los proveedores de servicios y el usuario pueden producir nuevos recursos personales con las herramientas adecuadas.
- **Gestor de recursos personales:** realiza distintas tareas de administración y control sobre los recursos personales disponibles y especialmente, sobre el intercambio de éstos dentro del modelo de negocio. De esta forma, este rol condiciona que los distintos actores puedan asumir los roles anteriormente definidos, otorgando y ejecutando permisos para la provisión, consumo o generación de recursos personales. Este rol lo asume principalmente el usuario, apoyándose en el uso de los servicios específicos del proveedor del marco de datos personales y a su vez, el propio marco de datos personales asume este rol, actuando en nombre del usuario y aplicando sus permisos establecidos.

Identificados los actores y roles del modelo de negocio, se puede establecer la relación entre los mismos. La siguiente tabla muestra un resumen de los actores y los roles que pueden asumir cada uno de ellos, conforme lo explicado para cada rol.

Tabla 4-1. Relaciones entre los actores y roles del modelo de negocio.

Rol \ Actor	Proveedor marco de datos personales	Proveedor de servicios	Usuario
Proveedor de recursos	Común	Común	Probable
Consumidor de recursos	Común	Común	Probable
Generador de recursos	Común	Probable	Probable
Gestor de recursos	Común	Probable	Común

Las relaciones entre los actores y roles se han clasificado como: común (cuando dentro del modelo de negocio lo normal es que un actor asuma este rol), probable (cuando el actor puede asumir dicho rol pero no es lo característico en el modelo de negocio) e improbable (cuando se considera que no puede darse el caso). Ninguna relación se ha establecido como improbable, ya que individualmente, todos los actores pueden asumir todos los roles.

En definitiva, consideramos que el proveedor del marco de datos personales asume todos los roles mencionados, obteniendo recursos personales de distintos proveedores, generando nuevos recursos personales y ofreciendo éstos a otros consumidores, todo lo anterior, aplicando las autorizaciones de los usuarios como gestor de recursos. Además, los proveedores de servicios finales actúan como proveedores o consumidores de recursos dentro del modelo de negocio, aunque es probable que individualmente asuman los roles de generador y gestor de recursos. Finalmente, el rol más característico de los usuarios es el de gestor de sus recursos personales pero con las herramientas y servicios adecuados, también podría asumir el resto de roles considerados. La siguiente figura muestra gráficamente las entidades participantes en el modelo de negocio, los roles que asumen y sus relaciones.

4.2.2 Propuestas de valor

El objetivo de una propuesta de valor es establecer qué producto y/o servicio se ofrece a los clientes de una empresa y otros grupos relacionados, identificando el valor que esto genera para ellos. Su finalidad es solucionar los problemas específicos de los clientes y satisfacer sus necesidades, aportando una serie de ventajas que otros negocios no ofrecen para permitir que los clientes se decanten por la solución propuesta.

Como se ha mencionado anteriormente, el modelo de negocio del marco de datos personales se basa en el patrón de plataforma multilateral y se dirige a dos segmentos del mercado principales: proveedores de servicios finales y usuarios. Para cada segmento, el marco de datos personales ofrece productos específicos relacionados con los datos personales, modelos de usuario y servicios de uso, generación y gestión de dichos recursos.

Basándonos en los actores y los roles identificados en la sección anterior, se establecen dos **propuestas de valor** principales del marco de datos personales:

1. El marco de datos personales ofrece a **proveedores de servicios finales**, que actúan como **consumidores de recursos personales**, datos personales y en concreto, modelos de usuario que se adaptan a sus necesidades de negocio específicas con conocimiento y consentimiento expreso del usuario. El valor percibido por los consumidores es el acceso a nuevos recursos personales de los usuarios desconocidos hasta el momento, eliminación de los riesgos asociados por la utilización de datos del usuario sin su consentimiento (por ejemplo, mala imagen de marca), simplificación y mejora del proceso de creación de recursos personales de los usuarios, facilidad de uso, personalización y adaptación de los recursos obtenidos, etc. Además, al participar el usuario en el modelo, los recursos personales que se obtienen son de mayor calidad, puesto que están verificados por ellos mismos. En definitiva, todo lo anterior se traduce en una reducción de los costes que debe ser determinante para fijar, si aplica, el precio que tienen que pagar por la propuesta de valor.
2. El marco de datos personales ofrece a sus **usuarios** distintos servicios para que actúen como **gestores de sus recursos personales** y así, puedan conocer qué recursos personales tienen disponibles y decidir quién accede a qué recursos y para qué. El valor principal que perciben los usuarios es el de control total sobre sus recursos

personales, lo cual está asociado a un aumento de transparencia. Además, los usuarios pueden percibir valor económico u otros incentivos (por ejemplo, descuentos, bonos u otros relacionados con el sentimiento de comunidad) por el hecho de permitir el acceso a sus recursos personales. También perciben valor relacionado con la facilidad de uso del servicio, la accesibilidad a recursos, personalización, automatización de los procesos, etc.

En principio, no sería necesario definir una propuesta de valor para los **proveedores de servicios finales** que actúan como **proveedores de recursos personales del marco de datos personales**, puesto que éstos previamente han evaluado y tomado la decisión en sus propios modelos de negocio de exponer sus recursos personales a terceras entidades. No obstante, conviene señalar que estos proveedores pueden exponer sus recursos para generar y favorecer la creación de nuevos recursos personales que, a su vez, pueden consumir y con ello, percibirían los valores descritos en la primera propuesta de valor. Adicionalmente, los proveedores de servicios finales podrían ofrecer en exclusiva al marco de datos personales sus recursos personales para favorecer la gestión de éstos a los usuarios. En sentido, dichos proveedores percibirían valor económico asociado a un incremento de la rentabilidad de los recursos personales que puede repercutirles directamente, una mejora de la imagen de marca, una disminución de costes asociada a la implementación de herramientas de gestión de la privacidad, etc.

Además de lo anterior, el **proveedor del marco de datos personales** también percibe a su vez valor derivado de los valores y beneficios obtenidos por el resto de actores y en particular, de las sinergias creadas entre los mismos. Así, el marco de datos personales obtendrá mayor beneficio cuantos más recursos personales permita el usuario incorporar de distintos proveedores, puesto que esto le permitirá aumentar las capacidades de generación de nuevos recursos personales y mejorar la oferta a los consumidores. Económicamente, los consumidores de recursos pueden pagar tanto al proveedor del marco de datos personales, como a los usuarios o los proveedores de recursos para atraer la incorporación y generación de recursos personales. Por tanto, se crean una serie de dependencias entre las ofertas de los consumidores y los permisos del usuario que determinan el valor global del marco de datos personales.

4.2.3 Segmentación del mercado

La segmentación del mercado consiste en la identificación y caracterización de los usuarios y/o clientes potenciales a los que va dirigido el marco de datos personales para poder ofrecer productos adaptados que garanticen su éxito. Dado que, hasta el momento, el marco de datos personales se ha definido como entidad genérica, sin especificar su ámbito de aplicación concreto, la segmentación del mercado no puede ser delimitada con precisión. Más adelante, en la sección 4.3 Particularización del modelo de negocio en distintos ámbitos de aplicación, se describirá la particularización del modelo de negocio del marco de datos personales en distintos ámbitos de aplicación (sector publicitario en redes sociales y el sector financiero) y se darán ejemplos de una segmentación de mercado más detallada.

En términos generales, el marco de datos personales se dirige a dos segmentos del mercado principales. Por una parte, pequeños o nuevos proveedores de servicios finales que requieren datos personales y en particular, modelos de usuario, para ofrecer mejores servicios y por otra parte, a los propios usuarios o personas físicas que se relacionan con el marco de datos personales para gestionar sus recursos personales. Éstos últimos se pueden caracterizar como personas extremadamente preocupadas por la utilización de sus datos personales o en el otro extremo, personas que quieren explotar y obtener beneficios tangibles de sus recursos personales sin preocuparse de ello.

4.2.4 Cadena de valor

La cadena de valor permite desagregar el modelo de negocio en distintas actividades que crean y distribuyen valor. Cualquier cadena de valor comienza con la entrada de determinados recursos y finaliza con la entrega del servicio y/o producto final a sus clientes y/o usuarios. Su definición permite determinar los activos necesarios para realizar la producción final y establecer la posición estratégica del negocio considerado en dicha cadena.

En el modelo de negocio del marco de datos personales, la cadena de valor se caracteriza por tener como recurso principal los datos personales de los usuarios y en particular, el nuevo conocimiento o modelos de usuario que se generan a partir de éstos para ofrecérselo a terceras entidades que crean nuevo valor. Esta cadena rompe con los enfoques tradicionales y actuales de los modelos de negocio próximos. Por una parte, modifica la logística de producción de modelos de usuario, considerándolo como un proceso distribuido en el que participan distintos actores. De esta forma, la cadena de valor comienza con el abastecimiento de datos personales de distintos proveedores de servicios, la generación de nuevo conocimiento y la exposición de los recursos anteriores a terceros proveedores de servicios que actúan como consumidores de los nuevos recursos generados para producir, a su vez, nuevo valor. Además, por otra parte, el usuario, que asume el rol de gestor de recursos, forma parte de la cadena de valor autorizando la realización de las distintas actividades de la cadena de valor. Así, el consentimiento del usuario es esencial para completar secuencialmente las actividades de la cadena de valor.

La Figura 4-2 refleja las actividades principales de la cadena de valor asociadas con los actores y roles participantes. En la parte superior se visualizan los recursos principales de la cadena de valor: datos personales que se transforman en modelos de usuario y que originan nuevo valor en los consumidores finales. En la parte inferior se ha representado la cadena de valor conforme a los distintos roles establecidos en el modelo de negocio (4.2.1 Actores, roles y relaciones).

Figura 4-2. Cadena de valor del modelo de negocio del marco de datos personales.

De la cadena de valor representada se observa que el marco de datos personales se sitúa como una entidad intermediaria entre distintos proveedores de servicios finales que asumen respectivamente los roles de proveedor y consumidor de recursos personales. El marco de datos personales lleva a cabo tres actividades principales: consumo, generación y provisión de recursos personales. La ejecución de cada una de ellas viene determinada por el gestor de recursos personales, bien el propio usuario o el marco de datos personales en nombre de éste. En definitiva, el marco de datos personales simplifica el proceso de modelado de usuario a los consumidores de recursos y permite a los usuarios modificar y actuar sobre la ejecución de cada una de las actividades consideradas, rompiendo con las cadenas de valor existentes actualmente.

4.2.5 Red de valor

La cadena de valor descrita anteriormente representa las actividades del modelo de negocio de forma secuencial y unidimensional. Dado que las relaciones que se pueden producir dentro de un modelo de negocio como el presentado son mucho más complejas que las representadas anteriormente, es necesario recurrir a otros elementos del modelo de negocio.

En este sentido, la red de valor es una evolución de la tradicional cadena de valor que considera todas las interconexiones que se establecen entre los actores participantes (o más específicamente, los roles que éstos asumen) y otras entidades para generar valor. Esto incluye la consideración de factores externos que pueden influir sobre el modelo de negocio (por ejemplo, aspectos legales o innovaciones tecnológicas) (Basole y Rouse, 2008) y la identificación del intercambio tanto de bienes tangibles (como servicios o ingresos) como intangibles (por ejemplo, conocimiento, emociones, etc.) entre los distintos participantes (Allee, 2010). En definitiva, supone un análisis complejo donde se consideran relaciones multidireccionales y simbióticas de negocio a negocio, negocio a consumidor o consumidor a consumidor (Basole y Rouse, 2008).

La siguiente figura representa gráficamente los aspectos más relevantes de la red de valor del marco de datos personales. Los nodos simbolizan los actores participantes: el marco de datos

personales (PeDF en la figura), el usuario y dos proveedores de servicios finales, aunque podrían ser más, que asumen el rol de proveedor y consumidor de recursos personales, respectivamente. Por otra parte, los arcos entre dichos nodos indican las distintas relaciones que se establecen entre los mismos. Con el objeto de simplificar la red de valor, se considerarán principalmente relaciones tangibles entre los participantes.

Figura 4-3. Red de valor para el marco de datos personales.

La red de valor del marco de datos personales sitúa a éste como punto central y alrededor el resto de actores que interactúan con él y a su vez, entre ellos, generando valor. Los usuarios y los proveedores de servicios finales que asumen el rol de consumidor de recursos personales son los que disparan la generación de valor en la red. Los primeros son los que aportan y facilitan el recurso principal de la red de valor, los datos personales, mientras que los segundos demandan al marco de datos personales el consumo de éstos y principalmente, del conocimiento generado y que no pueden alcanzar unilateralmente.

Simplificando todas las relaciones que pueden establecerse entre los usuarios y todos los proveedores de servicios (incluido el marco de datos personales) del modelo de negocio, se considera que entre éstos existen dos relaciones principales. Por una parte, los proveedores de servicios ofrecen servicios a los usuarios y éstos al utilizarlos aportan y crean datos personales que producen valor en cada uno de los proveedores. El uso que hagan los usuarios de estos servicios y en concreto, los permisos que establezcan desde el marco de datos personales para poder utilizar los datos que les refieren, es lo que determina el valor total de la red. En definitiva, si el usuario permite la utilización de sus datos personales, se generará valor en el resto de actores.

Por otra parte, los proveedores de servicios finales, que asumen el rol de consumidores de recursos personales, demandan al marco de datos personales distintos recursos personales que cubren sus necesidades de negocio específicas. Para responder a esta demanda, el marco de datos personales recurre a distintos facilitadores (proveedores de servicios que asumen el rol de proveedor de recursos personales) y obtiene los recursos demandados o los que permiten su generación. De nuevo, cuantos más datos personales disponga el marco de datos personales, mayor valor podrá generar y ofrecer a los consumidores de recursos.

Finalmente, cabe señalar que pueden establecerse distintas sinergias entre los proveedores de servicios y los usuarios que fomenten su participación en la red de valor, relacionadas, por ejemplo, con colaboraciones para exponer los recursos personales, ofertas para atraer la participación del usuario, etc. Todo lo anterior se establece mediante servicios ofrecidos por el marco de datos personales.

4.2.6 Mecanismos de generación de ingresos y otros incentivos

Hasta ahora, los elementos del modelo de negocio se han basado principalmente en la descripción de la participación y el valor relacionado con cada uno de los participantes. En este apartado, se define el flujo de los beneficios tangibles (generalmente, dinero) y los mecanismos que permiten su generación.

Los mecanismos de generación de ingresos del marco de datos personales pueden ser muy variables, dependiendo del ámbito considerado y el propietario del mismo, e incluso pueden no darse dentro del modelo de negocio. Por ejemplo, si se considera como propietario del marco de datos personales una institución pública que quiere fomentar el control de los recursos personales a los ciudadanos y la creación de nuevas empresas, el servicio puede ser un servicio público gratuito para todos los participantes. Lo mismo ocurriría en el caso de un organismo de investigación que quiere fomentar la experimentación o búsqueda de una solución. En estos casos, la participación de los actores viene dada por bienes intangibles relacionados con el sentimiento de comunidad o bienestar social, la motivación y otras emociones.

En términos generales, el flujo de ingresos (y en concreto, los pagos y cobros) dentro del modelo de negocio del marco de datos personales se basa en la redistribución de los beneficios tangibles que se da en modelos de negocio próximos donde existe un único beneficiario. De esta forma, se considera que los proveedores de servicios finales son los que pagan por los servicios ofrecidos por el marco de datos personales y este ingreso se distribuye entre el resto de participantes. La siguiente tabla resume los distintos pagos y cobros dentro del modelo de negocio:

Tabla 4-2. Resumen de los ingresos del modelo de negocio del marco de datos personales.

	COBRA	PAGA
Marco de datos personales	<ul style="list-style-type: none"> • Por servicios de mantenimiento de recursos personales • Por servicios de captación de nuevos recursos personales • Por servicios de generación de nuevos recursos personales y el uso de éstos 	<ul style="list-style-type: none"> • Comisiones a los usuarios por la utilización de sus recursos personales • Puede pagar comisiones a los proveedores de servicios cuyos recursos son solicitados
Consumidor	-	<ul style="list-style-type: none"> • Por la recuperación y uso de los recursos personales generados en el marco de datos personales. • Por la solicitud de incluir nuevos recursos personales en el marco de datos personales.
Proveedor	<ul style="list-style-type: none"> • Puede cobrar por exponer sus recursos personales 	<ul style="list-style-type: none"> • Comisiones a los usuarios por el uso de sus recursos personales • Puede contratar servicios de mantenimiento de recursos personales
Usuario	<ul style="list-style-type: none"> • Por aceptar el uso de sus datos personales de proveedores y consumidores de recursos personales 	-

En definitiva, los consumidores de recursos personales son lo que realizan los principales pagos en el modelo de negocio y su beneficio está relacionado con bienes intangibles como la satisfacción de cubrir una necesidad específica y los valores indicados en la propuesta de valor. Por su parte, el marco de datos personales recibe los pagos de los distintos proveedores de servicios que utilizan sus servicios y lo redistribuye entre el usuario y en ocasiones, los proveedores de recursos personales. Generalmente, esta redistribución se modela como un porcentaje de los ingresos recibidos. El usuario es el único que no tiene que pagar por ningún servicio del marco de datos personales.

Es importante señalar que se pueden considerar otros beneficios tangibles, no únicamente dinero dentro del modelo de negocio y especialmente, de cara al usuario. Por ejemplo, los consumidores de recursos personales pueden pagar al marco de datos personales por distintos servicios y ofrecer a los usuarios otros incentivos como, descuentos, cupones u otros programas de fidelización dentro del su servicio final ofrecido u otros asociados.

4.3 Particularización del modelo de negocio en distintos ámbitos de aplicación

En las secciones anteriores se ha descrito el modelo de negocio genérico del marco de datos personales, detallando los distintos elementos que lo conforman, la relaciones entre los

actores, los roles que éstos asumen y los valores tangibles e intangibles que se generan entre los mismos. Sin embargo, no se ha especificado ningún ámbito de aplicación concreto que permita comprender completamente las posibilidades del marco de datos personales. En el contexto de la tesis, se han realizado dos particularizaciones del modelo de negocio propuesto, una para el sector de publicidad en redes sociales y otra para el sector financiero, que han permitido validar soluciones específicas del enfoque global de la tesis. En los siguientes apartados se detallan dichas particularizaciones que suponen dos contribuciones originales de la tesis doctoral.

4.3.1 Modelo de negocio de intermediación de publicidad en redes sociales

4.3.1.1 Contexto y motivación

En 3.6.3.2 Modelo de negocio de publicidad basado en el uso de datos personales, se describió el modelo de negocio que distintos proveedores de servicios aplican para financiar la propuesta de valor gratuita ofrecida a sus usuarios. Dicho modelo es también característico de los servicios de redes sociales más populares de la actualidad, aunque generalmente lo combinan con otras propuestas de valor que permiten generar ingresos como la oferta de servicios avanzados de pago (*premium*), monedas virtuales, donaciones, juegos sociales, etc.

El origen de las redes sociales telemáticas no está claramente establecido pero puede remontarse a 1997 con la inauguración del sitio web *SixDegrees.com* y con otros servicios originarios (Ros-Martín, 2009). Desde entonces, han aparecido innumerables servicios de redes sociales (ver Figura 4-4), unos con más éxito que otros, hasta conformarse como plataformas centradas en el usuario que permiten la definición de distintas estructuras sociales y con ello, donde distintos grupos de personas se comunican, comparten contenidos (multimedia o pensamientos) y llevan a cabo distintas actividades relacionadas con el entretenimiento, la diversión o áreas profesionales para encontrar intereses comunes.

Figura 4-4. Lanzamiento de distintas redes sociales (1997-2010) (Fundación Telefónica, 2011).

En la actualidad, existe una gran variedad de redes sociales y las posibilidades que éstas ofrecen son múltiples. Por una parte, podemos encontrar redes sociales horizontales, cuyo objetivo principal es crear una gran masa de usuarios que se relacionan entre sí sin que exista una temática concreta (por ejemplo, Facebook, Google+ o Tuenti), y por otra parte, redes sociales verticales que se especializan en una temática de interés para un colectivo concreto (por ejemplo, LinkedIn o Twitter). De esta forma, se pueden encontrar clasificaciones de las redes sociales por la temática que tratan (profesionales, aficiones, viajes, idiomas, etc.), por las actividades que se realizan (juegos, geolocalización, compartición de imágenes, etc.) o por los contenidos que se comparten (reflexiones con caracteres limitados, fotos, música, vídeos, lecturas, etc.). En términos generales, las redes sociales permiten a los usuarios crearse un espacio personal, denominado generalmente como perfil de usuario, donde comparten distintos contenidos con el resto de usuarios y que favorece la comunicación y el encuentro de intereses comunes entre los mismos.

Los servicios de redes sociales se han convertido en los principales motores de la evolución de Internet en los últimos años, transformando la forma en la que se prestan hasta otros servicios que aparentemente no están relacionados. Su influencia es tal, que se han configurado como un canal idóneo para que otros proveedores de servicios puedan alcanzar a sus clientes, dando lugar al denominado marketing de medios sociales (Weinberg, 2009). Este nuevo canal facilita que anunciantes y usuarios se encuentren en un nuevo contexto, interactúen y creen nuevos vínculos, reduciendo los costes que se producirían en medios tradicionales. Sin embargo, requiere que los proveedores de servicios adquieran nuevas habilidades para alcanzar a sus clientes potenciales en el nuevo contexto.

Distintas redes sociales, como Facebook o Twitter, ofrecen a proveedores de servicios la posibilidad de posicionar dentro de sus portales anuncios publicitarios dirigidos, apoyándose para ello en el número elevado de usuarios a los que pueden alcanzar y en el conocimiento

que poseen de éstos. No obstante, aunque aparentemente las redes sociales pueden suponer un medio de publicidad idóneo, el modelo de negocio basado en publicidad para las redes sociales y especialmente en el momento en el que se realizó el trabajo de investigación para esta tesis doctoral, no había terminado de desplegarse (Carlsson, 2010) y éstas constantemente están buscando nuevas soluciones y diversificando su negocio para obtener nuevos ingresos.

En los últimos años, las redes sociales han optimizado sus mecanismos de publicidad con nuevas soluciones, como por ejemplo, la promoción de publicaciones de Facebook (2015) o las cuentas y *tweets* (mensajes dentro de la red social) promocionados de Twitter (2015), que se basan en la integración de los anuncios como si fueran un contenido más de los amigos de un usuario de la red social. Sin embargo, estos contenidos son emitidos directamente por los proveedores de servicios y difundidos por las redes sociales entre el público objetivo. Respecto a lo anterior, también se ha desarrollado un nuevo modelo de negocio paralelo que se basa en el marketing de influencia y que consiste en que distintos proveedores de servicios patrocinan o pagan a personas relevantes de un sector (generalmente, famosos) por la publicación de contenidos relacionados con sus productos o servicios (Augure, 2014).

El modelo de negocio presentado en este apartado supone una combinación de las soluciones que actualmente se están aplicando en el contexto de redes sociales y se propuso en el proyecto *Segur@* en el año 2010, fruto de la colaboración con la empresa Ericsson España, cuando no existían ninguna de las aproximaciones mencionadas. Hasta donde alcanza nuestro conocimiento, no existe actualmente un modelo de negocio como el presentado y en el momento en el que se realizó la investigación, únicamente se detectó una aproximación similar en (Lim, 2005).

4.3.1.2 Descripción del modelo de negocio

El modelo de negocio propuesto se basa en incorporar dentro de la cadena de valor del modelo de negocio publicitario de redes sociales a los usuarios para que éstos emitan distintos anuncios dentro de sus perfiles de usuario y así, lleguen a sus amigos, compañeros o familiares. Esto permite aprovechar los valores asociados al marketing boca a boca electrónico, como por ejemplo, la confianza que los usuarios depositan en sus allegados sobre sus opiniones comerciales, el efecto viral, la interactividad con los anuncios, la reducción de costes, etc. (Jansen, et al., 2009).

En este contexto, se considera como anuncio a cualquier contenido (texto, imagen, vídeo, audio, etc.) que refleja la relación entre un usuario y un proveedor de servicios (como por ejemplo, la satisfacción por una compra, la opinión sobre el servicio prestado, el diseño del sitio web, etc.) y que puede utilizarse con fines publicitarios. Además, como dicho contenido está emitido por un usuario y hace referencia a distintos aspectos de éste, se considera que un anuncio es un dato personal del usuario.

Dadas las consideraciones anteriores, proponemos la especialización del proveedor del marco de datos personales como un intermediario de anuncios publicitarios que ha sido denominado como *Online Social Advertisement Mediator* (OSAM).

En el modelo de negocio del OSAM (ver Figura 4-5) participan cuatro actores principales: redes sociales, proveedores de servicios, usuarios y el propio OSAM.

- Las **redes sociales** son proveedores de servicios especializados en ofrecer a los usuarios un medio de comunicación y compartición de contenidos con otros usuarios y además, ofrecen distintos recursos a terceras entidades como datos personales o servicios avanzados relacionados con la publicación de contenidos, la autenticación en servicios externos, el desarrollado de aplicaciones en sus plataformas, etc. Como se señaló anteriormente, las redes sociales cuentan con los recursos necesarios para ser un canal publicitario idóneo para terceros proveedores.
- Los **proveedores de servicios** son entidades que se relacionan con los usuarios ofreciendo servicios finales de distintos ámbitos y que quieren publicitarse para alcanzar a nuevos clientes potenciales en una o más redes sociales de los usuarios.
- Los **usuarios** son las personas que utilizan tanto los servicios ofrecidos por distintas redes sociales, como los de distintos proveedores de servicios con los que interactúan, incluido el propio OSAM, y además, adquiere un nuevo rol dentro del modelo de negocio convirtiéndose en emisor de anuncios publicitarios dentro de sus redes sociales.
- El **OSAM** actúa como una entidad intermediaria entre el resto de actores y ofrece y aprovecha recursos y servicios específicos a cada uno de ellos. En particular, incorpora los servicios necesarios para que los proveedores de servicios emitan a través del usuario anuncios dirigidos en distintas redes sociales y las herramientas necesarias para que los usuarios puedan controlar la publicación en su nombre dichos anuncios. Todo ello aprovechando los recursos y servicios ofrecidos por distintas redes sociales.

Figura 4-5. Visión global del modelo de negocio del OSAM.

Al igual que el modelo de negocio de referencia establecido para el marco de datos personales, el OSAM aplica el patrón de modelo de negocio de plataforma multilateral y ofrece una propuesta de valor a cada uno de los actores participantes:

- Se dirige a **proveedores de servicios** que quieren anunciarse y les ofrece un nuevo procedimiento simplificado que permite alcanzar una gran masa de clientes potenciales en distintos canales (redes sociales), dirigiendo la publicidad a las necesidades específicas de los proveedores y que están relacionadas con las características particulares de los clientes potenciales. El valor que perciben los proveedores es que pueden publicitarse en múltiples redes sociales con un único procedimiento, reduciendo carga de trabajo y costes. Además, los anuncios emitidos son más efectivos al realizarlos los usuarios y están dirigidos a clientes potenciales conforme a unas condiciones y necesidades de negocio específicas que previamente los proveedores de servicios han establecido en el OSAM, lo que cual es altamente valorado por dichos proveedores. En definitiva, la propuesta de valor supone una mejora del proceso publicitario que cubre las necesidades específicas de los proveedores de servicios.
- Incluye las herramientas necesarias para que los **usuarios** puedan participar en el modelo de negocio publicitario y emitan anuncios en su nombre en sus distintas redes sociales, manteniendo el control sobre sus recursos y respetando su privacidad frente a los proveedores de servicios. A cambio, los usuarios obtienen ingresos, descuentos comerciales y otros incentivos ofrecidos por los proveedores de servicios pero gestionados a través del OSAM. También, los usuarios perciben una automatización de la generación de anuncios que facilita la tarea de emisión y divulgación de sus anuncios y el control de todo el proceso basado en principios de transparencia.

Además de lo anterior, el OSAM establece relaciones de confianza y acuerdos comerciales con distintas **redes sociales** para que éstas se constituyan como canales publicitarios dentro del modelo de negocio. Esto supone que las redes sociales garantizan la prestación de servicios, como la publicación de contenidos desde entidades externas, y la recuperación de los datos personales requeridos para que el OSAM lleve a cabo la publicidad dirigida. A cambio de lo anterior, las redes sociales pueden complementar o incluso, externalizar por completo, su modelo de negocio basado en publicidad y con ello, percibir valor económico con los nuevos ingresos generados en el modelo de negocio. Por tanto, el valor que perciben las redes sociales es una simplificación y mejora en los procedimientos publicitarios que implica una reducción de costes y un incremento de ingresos y por otra parte, un aumento del uso de sus servicios, ya que se generan nuevas publicaciones de los usuarios (los anuncios que emiten desde proveedores externos) que a su vez pueden generar nuevos comentarios de otros usuarios de su red social, y con ello, del tráfico generado.

En definitiva, el OSAM supone un punto de encuentro entre redes sociales que quieren potenciar la utilización de sus servicios o sus modelos de negocio basados en publicidad y proveedores de servicios que quieren mejorar su impacto publicitario. Para ello, el OSAM incluye los servicios y herramientas necesarias que permiten a los usuarios participar activamente dentro del modelo de negocio publicitario y convierte a éstos en emisores de

anuncios que se divulgan dentro de sus redes sociales, lo cual aumenta la efectividad de los mismos.

La siguiente figura representa la cadena de valor del modelo de negocio del OSAM, suponiendo que se han preestablecido unos acuerdos de garantía de prestación de servicios en determinadas redes sociales y las condiciones publicitarias del proveedor de servicios que quiere anunciarse.

Figura 4-6. Cadena de valor del modelo de negocio del OSAM.

La cadena de valor del OSAM es muy similar a la propuesta para el marco de datos personales, puesto que éste se sitúa entre proveedores de servicios y redes sociales (proveedores de servicios especializados). Sin embargo, para comprender el modelo de negocio del OSAM, se han detallado las actividades concretas relacionadas con la participación del usuario sobre dicha cadena.

De esta forma, la cadena comienza cuando un usuario que interactúa con un proveedor de servicios específico, realizando distintas acciones (por ejemplo, una compra o una consulta de un producto), recibe una oferta del proveedor de servicios para que emita un anuncio en sus redes sociales que refleje la relación entre ambos. En dicha oferta, el usuario visualiza las condiciones (por ejemplo el contenido del anuncio o parte de éste, las redes sociales para su divulgación, etc.) y los incentivos que puede obtener por la emisión del anuncio. Si, finalmente, el usuario acepta la oferta de publicar un anuncio en su nombre, el proveedor de servicios se dirige al OSAM para que publique el anuncio del usuario conforme a las condiciones que previamente se han establecido. En este punto, el OSAM debe elaborar los anuncios para que sean los más dirigidos posibles dentro de la red social del usuario y se cubran las necesidades de negocio preestablecidas por los proveedores de servicios. Esto puede requerir la recuperación de datos personales de los usuarios en distintas redes sociales para permitir dirigir la publicidad correctamente. Una vez que ha elaborado el anuncio final, el OSAM se lo presenta al usuario para su aceptación y esto genera la divulgación en las redes sociales establecidas en las condiciones.

Como se explicó en la sección anterior, la cadena de valor es una representación secuencial de las actividades que se llevan a cabo en el modelo de negocio, sin embargo, no considera otras relaciones que generan valor dentro del mismo. En términos generales, las relaciones entre los

distintos actores que crean valor dentro del modelo de negocio ya se ha reflejado en la Figura 4-5 que ofrece una visión global del modelo de negocio.

Los usuarios generan valor al utilizar los distintos servicios ofrecidos por las redes sociales, los proveedores de servicios y el OSAM. A cuantas más redes sociales pertenezca el usuario y más activo esté dentro de las mismas, mayor valor puede generar y ofrecer el OSAM a los distintos proveedores de servicios que quieran publicitarse, puesto que la masa de clientes potenciales será mayor. Por su parte, el OSAM se dirige a múltiples proveedores de servicios y cuanto mayor sea el número de éstos, mayor valor generará en las redes sociales asociadas con el OSAM, puesto que podrán emitir más anuncios y esto generará más tráfico. En definitiva, el OSAM se dirige a tres segmentos del mercado interdependientes y el valor que se genera para cada uno de ellos depende del resto de segmentos, aumentando a media que lo hace el número de cada uno de ellos.

Por otra parte, el OSAM puede actuar como un agregador de segmentos de clientes y/o usuarios situados en la larga cola, ofreciendo a proveedores de servicios publicidad muy dirigida en redes sociales especializadas que reúnen usuarios con características muy concretas y dispares. Finalmente, dado que el OSAM posee datos personales de todas las relaciones existentes, puede generar nuevo valor para ofrecer información a los proveedores de servicios.

Para finalizar, cabe señalar que los mecanismos de generación de ingresos del modelo de negocio presentado provienen en su mayoría de los proveedores de servicios que quieren anunciarse en distintas redes sociales. Éstos pagan al OSAM una cantidad que viene determinada por las condiciones que establecen en el acuerdo de publicación de anuncios, por ejemplo, número de redes sociales, segmentación de los usuarios que reciben los anuncios, etc. El OSAM distribuye el pago recibido entre el resto de actores, redes sociales y principalmente, el usuario, aunque este último también puede percibir otros incentivos como descuentos, bonificaciones u otras ventajas aplicables en los proveedores de servicios anunciantes. Además, dependiendo del acuerdo llegado entre las redes sociales y el OSAM, éstas pueden pagar también por participar dentro del modelo de negocio y de nuevo, la cantidad se repartiría con el usuario aplicando un porcentaje proporcional por cada uno de los anuncios que emitan.

4.3.2 Modelo de negocio para la promoción de nuevos servicios financieros

4.3.2.1 Contexto y motivación

El sector financiero es uno de los sectores más tradicionales y maduros donde distintas empresas y entidades, como, por ejemplo, bancos y otras instituciones financieras, mercados de valores y aseguradoras, ofrecen servicios de intermediación para generar valor a través del dinero.

Al igual que ha sucedido en otros mercados como el del transporte, el alojamiento o el consumo audiovisual, la irrupción de la tecnología está transformando los servicios y procedimientos habituales del sector financiero y amenazan a los actores tradicionales,

aunque éstos habían conseguido posponer su impacto dado su marcado acento tradicional y la dura regulación sobre su actividad.

De esta forma, el sector financiero ha sido testigo de cómo en los últimos años ha aparecido una serie de nuevos actores, distintas compañías del sector de las tecnologías de la información y nuevos proveedores de servicios financieros, que han aprovechado las oportunidades que brindan la tecnología para ofrecer nuevos servicios financieros centrados en el usuario, soluciones innovadoras que permiten la desintermediación de las actividades tradicionales del sector y en definitiva, formas radicalmente distintas de hacer negocio. Tal es el impacto que estos actores conforman un nuevo sector denominado como FinTech (acrónimo para denotar tecnología financiera, del inglés, Financial Technology). Éste presenta índices de crecimiento positivos con una media del 26% desde el año 2008 y previsiones favorables para los años venideros (Gach y Gotsch, 2014).

En el caso particular de los bancos, tradicionalmente, éstos han sido pioneros en la adopción de nuevas tecnologías y han ido evolucionado para ofrecer a sus clientes servicios telemáticos que permiten contratar nuevos productos y servicios financieros, realizar y consultar transacciones bancarias, domiciliar recibos y otros pagos y, en general, llevar a cabo todas las operaciones de la banca tradicional. Sin embargo, a pesar de los avances, la aplicación y las investigaciones relacionadas con la tecnología por parte de los bancos han estado principalmente enfocadas a la optimización de sus procesos internos o a la adaptación por exigencias de entidades externas, como por ejemplo, organismos de regulación. De esta forma, raramente los usuarios han percibido las innovaciones realizadas por los bancos (Barri, et al., 2014).

Los nuevos actores han sabido aprovechar las carencias y huecos de los bancos y otras entidades financieras para proponer nuevos modelos de negocio que les permiten reaccionar rápidamente a las necesidades de los usuarios y el contexto y adaptarse a los hábitos de consumo. Todo ello ha provocado que se posicionen como alternativas modernas frente a los bancos y que los usuarios depositen su confianza en los nuevos actores (Barri, et al., 2014).

Por ejemplo, los servicios de pago telemáticos se han convertido en una de las áreas más competitivas y múltiples compañías tecnológicas, como Paypal, Google o Apple, han irrumpido para ofrecer soluciones innovadoras. Principalmente, estos servicios han avanzado en dos direcciones. Por una parte, se han desarrollado multitud de soluciones que incorporadas en los teléfonos móviles de los usuarios facilitan la realización de pagos en contextos físicos y por otra parte, se han desarrollado nuevos mecanismos que facilitan la realización de pagos en contextos telemáticos, bien entre una persona y una organización, o bien para el intercambio directo de dinero entre personas. Estos últimos añaden grandes ventajas a los usuarios como, por ejemplo, la devolución de dinero, flexibilidad de horario, mecanismos de seguridad sofisticados pero sencillos para los usuarios, posibilidad de anonimato en las transacciones, la desintermediación aparente, etc. (Kim, et al., 2010).

Además del ejemplo anterior, los nuevos proveedores de servicios ofrecen servicios que permiten llevar a cabo *crowdfunding*, fomentar los préstamos entre personas (como los *lending clubs*), obtener recomendaciones de nuevas inversiones, agregar información

financiera para gestionar las finanzas personales, comparar productos u obtener recomendaciones de productos financieros que se adapten a las necesidades de los usuarios, comprar o vender monedas virtuales e incluso, divisas, etc. En definitiva, nuevas ofertas y servicios que hasta ahora no estaban al alcance de todos.

Aunque aparentemente la irrupción de los actores anteriores puede suponer una serie de amenazas para la industria financiera, ésta también representa una oportunidad para que los actores tradicionales se posicionen en el nuevo contexto. En este sentido, los actores tradicionales del sector cuentan con un gran número de ventajas competitivas inherentes como, por ejemplo, una gran masa de clientes, datos sobre sus transacciones bancarias, la infraestructura necesaria para la realización de pagos y préstamos de forma segura, etc. (Busch y Moreno, 2014) y (Open Data Institute y Fingleton Associates, 2014).

Conscientes de esta situación, el Banco Santander a través del Center for Open Middleware lanzó en 2014 un proyecto piloto denominado *POSDATA – Personal and Social Data Analysis* que tenía entre sus objetivos específicos el análisis de nuevos modelos de negocio en torno a los datos personales de los usuarios. En el siguiente apartado se describe el modelo de negocio propuesto en el contexto del proyecto y que ha sido validado en el enfoque global de la tesis doctoral.

4.3.2.2 Descripción del modelo de negocio

Dada la necesidad de los actores tradicionales del sector financiero por posicionarse dentro del contexto anteriormente descrito, se propone la especialización del marco de datos personales como una **bróker financiero** que permita establecer un nuevo ecosistema en torno a los datos personales y financieros de los usuarios para la fomentar la creación de nuevos servicios y con ello, la generación de valor que repercuta sobre todos los actores participantes.

El modelo de negocio del bróker financiero (ver Figura 4-7) se establece por la participación de cinco actores principales: proveedores de servicios financieros tradiciones (como bancos o aseguradoras), fuentes de datos externas disponibles (por ejemplo, redes sociales o plataformas de datos abiertas), los usuarios, nuevos proveedores de servicios (principalmente del ámbito financiero) y el proveedor del bróker financiero.

Figura 4-7. Visión global del modelo de negocio del bróker financiero.

Al igual que en el modelo de negocio de referencia del marco de datos personales, el bróker financiero se apoya en la utilización de recursos personales de distintos proveedores para generar nuevos recursos personales (generalmente, modelos de usuario) que ofrece a terceras entidades que actúan como consumidores de recursos, todo ello con la supervisión y control de los usuarios.

En particular, en el modelo de negocio del bróker financiero se considera dos tipos de proveedores de recursos personales. Por una parte, **proveedores de servicios financieros tradicionales**, como, por ejemplo, bancos, mercados de valores o aseguradoras, que poseen datos financieros y otros datos personales de los usuarios, que cuentan con una gran masa de clientes a los que ofrecen servicios tradicionales y habituales del sector financiero. Por otra parte, se consideran como proveedores de recursos personales a otras **fuentes de datos personales** existentes, por ejemplo, plataformas de datos abiertos, redes sociales, etc., que poseen recursos, generalmente de otros ámbitos fuera del financiero, y que pueden utilizarse para la generación o enriquecimiento de modelos de usuario del ámbito financiero.

Además, de lo anterior, participan en el modelo **nuevos proveedores de servicios** que actúan como consumidores de modelos de usuario del bróker financiero. Estos proveedores son nuevos entrantes del sector financiero, aunque puede extenderse a otros sectores, que requieren recursos personales de los usuarios para ofrecer o mejorar sus servicios finales.

Finalmente, los **usuarios** son las personas físicas que disfrutan de los servicios ofrecidos por los proveedores de servicios tradicionales, otras fuentes de datos, el propio bróker financiero y los nuevos proveedores de servicios y en el modelo de negocio, poseen las herramientas necesarias para gestionar el flujo de su información para decidir quién accede a sus recursos personales y para qué a cambio de distintos incentivos.

El **proveedor del bróker financiero** incluye todos los servicios necesarios para la participación del resto de actores dentro del modelo de negocio. Al igual que el marco de datos personales, aplica el patrón de modelo de negocio de **plataforma multilateral** y se dirige a distintos segmentos del mercado ofreciendo una propuesta de valor a cada uno de éstos. En particular, las propuestas de valor que ofrece son:

- Se dirige a **proveedores de servicios financieros tradicionales** para que actúen como proveedores de recursos personales dentro del modelo de negocio. Esto permite la desagregación del modelo de negocio tradicional de las instituciones financieras que puede percibirse como una oportunidad, en concreto, en aspectos relacionados con la innovación de nuevos productos y servicios. De esta forma, los proveedores de servicios financieros tradicionales valoran la desagregación de su modelo de negocio, puesto que les permite externalizar la creación de nuevos servicios financieros que se adapten a las necesidades de los usuarios y el contexto sin entrar en conflicto con sus actividades habituales y que repercuten positivamente sobre ellos mismos. Así, principalmente, los proveedores de servicios financieros perciben una reducción de costes para explorar nuevas oportunidades en torno al uso, gestión y aprovechamiento de sus recursos personales y todo ello se traduce en una mejora de la imagen de marca, puesto que se posicionan como proveedores de servicios abiertos a los nuevos tiempos, innovadores y en definitiva, modernos frente a los usuarios.
- Se dirige a **nuevos proveedores de servicios** que requieren recursos personales y en particular, modelos de usuario relacionados con características financieras de los usuarios, para prestar sus servicios. El valor que perciben estos nuevos proveedores son el acceso a recursos especializados de calidad que cubren sus necesidades de negocio específicas y les permite prestar sus servicios rompiendo con las barreras de entrada del sector: el acceso a recursos personales. Además, perciben una reducción de costes y la simplificación del proceso de modelado de usuario y otros valores relacionados con la personalización y la facilidad de uso.
- Se dirige a los **usuarios** para ofrecerles los servicios necesarios que les permite participar en el modelo de negocio de sus datos personales y obtener distintos beneficios económicos u otros incentivos comerciales. De esta forma, el valor que perciben los usuarios es principalmente económico pero también valoran el control sobre el uso de sus datos personales, las ofertas de nuevos servicios que pueden crear los nuevos proveedores de servicios y valores relacionados con la facilidad de uso del servicio, la automatización de los procesos, la personalización, etc.

En principio el bróker financiero no se dirige a otras fuentes de datos personales que ofrecen sus recursos a terceras entidades pero se pueden establecer acuerdos de colaboración para garantizar la disponibilidad de distintos recursos que permitan enriquecer los datos financieros ofrecidos por los proveedores tradicionales. En términos generales, las fuentes de datos personales perciben valores relacionados con los ecosistemas de plataformas de datos abiertas, donde la creación de nuevo valor con sus recursos puede repercutir directamente sobre éstas.

Por otra parte, en lo que refiere a la cadena de valor del bróker financiero, ésta es muy similar a la ofrecida por el marco de datos personales (4.2.4 Cadena de valor). De esta forma, como se puede deducir de las explicaciones anteriores, el bróker financiero se sitúa entre distintos proveedores (proveedores de servicios tradicionales y nuevos proveedores de servicios finales) y realiza tres actividades principales: **recolección** de datos financieros de los usuarios (y otros datos personales de fuentes de datos disponibles), **generación** de modelos de usuario con características financieras y **provisión** del nuevo conocimiento generado a terceras entidades. Además, cada una de las actividades está fijada por las **políticas** establecidas por el usuario que determinan la ejecución de la cadena.

De igual forma, la red de valor es análoga a la establecida para el marco de datos personales (4.2.5 Red de valor), especializando los proveedores de servicios participantes a actividades del ámbito financiero. A diferencia de la red de valor de referencia, el valor que se crea entre los distintos actores participantes puede no estar determinado por el número de proveedores de servicios financieros tradicionales participantes, puesto que esto depende específicamente del volumen de datos financieros que posean dichos proveedores y con ello, del número de sus clientes. Es decir, una única entidad financiera con una gran masa de clientes, como, por ejemplo, el Banco Santander, puede generar suficiente valor al resto de los actores participantes, puesto que los datos financieros que estos poseen pueden cubrir las necesidades específicas de un gran número de nuevos proveedores de servicios.

Finalmente, en cuanto con los mecanismos de generación de ingresos, éstos provienen principalmente de los nuevos proveedores de servicios a los que se les ofrece distintos planes de financiación dependiendo de sus características particulares y de las condiciones preestablecidas por los proveedores de servicios financieros tradicionales. Por ejemplo, los nuevos proveedores podrán aportar una cantidad determinada al bróker financiero por obtener determinados recursos o se podrán realizar pagos determinados por un porcentaje sobre los beneficios obtenidos por los nuevos entrantes. Independientemente del mecanismo de pago, los ingresos obtenidos por el bróker financiero se redistribuyen con el usuario y en ocasiones, con los proveedores de servicios financieros tradicionales. También se considera la oferta de distintos incentivos comerciales a los usuarios en lugar de o complementando el reparto de dinero.

4.4 Resumen y contribuciones del capítulo

En este capítulo se ha presentado un nuevo modelo de negocio que permite avanzar hacia un ecosistema de intercambio de datos personales centrado en el usuario aprovechando las oportunidades del contexto actual. En primer lugar, se ha llevado a cabo un análisis de los modelos de negocio que actualmente se basan en el uso y la compartición de datos personales entre distintas entidades, destacando las fortalezas y debilidades de cada uno de ellos, lo que ha permitido posicionar el nuevo modelo de negocio considerado en los huecos existentes. Seguidamente, para facilitar su comprensión, se ha aplicado un enfoque que va de lo abstracto a lo particular para describir el nuevo modelo de negocio. De esta forma, se comenzó con una descripción general de los distintos elementos del modelo de negocio que permite obtener una visión global de las posibilidades de éste y se continuó con la explicación del modelo de

negocio en distintos ámbitos de aplicación del contexto privado. En particular, se ha descrito un nuevo modelo de negocio para el sector publicitario y otro para el sector financiero.

En consecuencia, las contribuciones originales que se pueden extraer de este capítulo son:

- Análisis crítico de los distintos modelos de negocio que existen actualmente respecto al uso y la compartición de datos personales con terceras entidades para detectar nuevas oportunidades de negocio.
- Propuesta de un nuevo modelo de negocio de referencia que aprovecha las oportunidades detectadas, fomentando como propuestas de valor de éste las debilidades de los modelos de negocio actuales.
- Particularización del modelo de negocio en el sector publicitario, lo cual incluye un análisis específico del contexto publicitario y la detección de huecos existentes para la propuesta de un nuevo modelo de negocio basado en la intermediación de información respetando la privacidad del usuario.
- Particularización del modelo de negocio en el sector financiero, lo cual implica un estudio del sector, detectando las últimas tendencias y la evolución del mercado y con ello, la propuesta de un nuevo modelo de negocio que permite posicionarse a los proveedores tradicionales para la promoción de nuevos servicios financieros.

A diferencia de como se han presentado las contribuciones, la investigación se ha realizado siguiendo una metodología inductiva que permite la observación y abstracción de resultados en distintos ámbitos para su generalización posterior. De esta forma, los primeros trabajos relacionados con el modelo de negocio se incubaron en los proyectos Consorcios Estratégicos Nacionales en Investigación Técnica (CENIT) Segur@ y Buscamedia (Buscamedia, 2015), donde se validaron distintos modelos de negocio basados en la intermediación de información en el ámbito de la publicidad y los contenidos audiovisuales y posteriormente, la aplicación al sector financiero y la abstracción del modelo del negocio que se llevó a cabo en el proyecto POSDATA del Center for Open Middleware.

Los resultados más relevantes de las contribuciones incluidas en el capítulo son:

- Solicitud y posterior aceptación de una patente que actualmente está en explotación por parte de la empresa Ericsson. Ésta define los métodos para permitir la distribución de información entre las entidades del modelo de negocio de publicidad en redes sociales presentado (Monjas et al., 2012).
- Elaboración y aceptación de publicación del artículo titulado *A Personal Data Framework for Distributed User Modelling in New Service Ecosystems* (San Miguel, et al., 2015e) en la revista IEEE América Latina indexada en JCR SCI.
- Publicación del artículo titulado *A Personal Data Framework for Exchanging Knowledge about Users in New Financial Services* (San Miguel, et al., 2015a) que fue presentado en el *1st International Workshop on Personalization & Recommender Systems in Financial Services (FINREC)*. Éste describe el modelo de negocio en el sector

financiero para la promoción de nuevos servicios y describen nuevos servicios que han sido desarrollados en el contexto de esta investigación.

- Publicación del artículo titulado *Extracting Knowledge through Friendsourcing: Application to new Financial Services* que fue presentado en el VII Congreso Iberoamericano de Telemática (CITA2015) y que fue seleccionado como uno de los mejores artículos del congreso. Este artículo se centra en el modelo de negocio del sector financiero profundizando en nuevos servicios de micromecenazgo (*crowdfunding*) que se pueden originar en torno al marco de datos personales.

En definitiva, la experiencia obtenida en relación a los modelos de negocio originó que se abrieran nuevas líneas de investigación y validaciones técnicas que se presentarán en el resto de capítulos. Así, en el siguiente capítulo se detallará una especificación formal para el marco de datos personales que permitirá establecer una arquitectura de referencia que se sustenta en el modelo de negocio presentado. Dicha arquitectura será la base sobre la cual se construirá el resto de contribuciones de la tesis doctoral incluidas en los capítulos posteriores.

Capítulo 5.

Especificación y diseño de un marco de datos personales de referencia

En el capítulo anterior, se describió el modelo de negocio de referencia para un ecosistema de intercambio de recursos personales centrado en el usuario. De este modo, se presentaron los detalles de los distintos elementos de éste y su particularización en el ámbito de la publicidad para la intermediación de anuncios en redes sociales y en el sector financiero para la promoción de nuevos servicios. Todo lo anterior permite obtener una visión global del sistema y del escenario de referencia de la tesis doctoral. No obstante, aunque se puedan intuir algunos de los requisitos necesarios para la realización del marco de datos personales, conviene formalizar una especificación y posteriormente, el diseño, que permita establecer con precisión un marco de datos personales de referencia.

El marco de datos personales se sitúa en torno a distintos proveedores de servicios que proporcionan y/o consumen recursos personales y el usuario que gestiona el flujo de los datos personales que le refieren. Para su especificación, los proveedores de recursos personales serán denominados genéricamente como fuentes de datos y éstas podrán ser públicas o privadas, dependiendo de las restricciones de acceso que establezcan sobre sus recursos (ver 3.3 Mecanismos de recuperación de datos personales almacenados en fuentes de datos externas). Por otra parte, los proveedores de servicios que requieren recursos personales del marco de datos personales se denominarán como consumidores y en particular, éstos serán consumidores de modelos de usuario dada la novedad que añade este enfoque al modelo de negocio.

Dadas las consideraciones anteriores, la siguiente figura representa la visión global de un marco de datos personales de referencia donde se han reflejado distintas fuentes de datos (públicas y privadas) que proveen de datos personales, consumidores que obtienen modelos de usuario para cubrir sus necesidades de negocio específicas y al usuario que ejerce tareas de administración y control para gestionar la privacidad sobre los recursos personales que le competen directamente.

Figura 5-1. Visión global para la especificación de un marco de datos personales de referencia.

La organización del capítulo es la siguiente: en la sección 5.1 se formalizará la especificación de un marco de datos personales, considerando no sólo los servicios que éste debe llevar a cabo, sino también los principios en los que éste debe estar fundamentado. Lo anterior permitirá realizar el diseño de una arquitectura de referencia que se detalla en la sección 5.2 junto con la descripción de cada uno de sus módulos y componentes. Por último, la sección 5.3 incluye el resumen y las contribuciones originales del capítulo, así como una descripción del enfoque global de validación realizado sobre la arquitectura de referencia y su distribución en el resto del documento.

5.1 Especificación del marco de datos personales

La especificación de un sistema software permite establecer lo que debe realizar dicho sistema y las restricciones en sus operaciones (Sommerville, 2006). Lo anterior implica la identificación de requisitos funcionales y no funcionales. Los primeros describen qué debe hacer el sistema de forma concreta y completa y así, definen los servicios y funciones del mismo, mientras que los requisitos no funcionales limitan las características globales del sistema y con ello, determinan su correcto funcionamiento.

En esta sección se incluye la especificación de un marco de datos personales de referencia siguiendo un enfoque genérico y por lo tanto, sin hacer alusión a ninguna tecnología específica. En el primer apartado (5.1.1), se presenta una descripción funcional de alto nivel y seguidamente, en 5.1.2 se incluye una especificación complementaria del sistema que refleja los requisitos no funcionales del mismo.

5.1.1 Descripción funcional de alto nivel mediante casos de uso

Los casos de uso son una técnica para la especificación de requisitos funcionales de un sistema que muestran a través de distintos diagramas y descripciones en lenguaje natural los procesos que deben llevarse a cabo para proporcionar un resultado observable y relevante para un determinado actor (Larman, 2001). Éstos se expresan desde el punto de vista del actor principal que inicializa el caso de uso pero pueden requerir la participación de otros actores (denominados como de soporte o secundarios) para llevar a cabo la ejecución o activación de otros casos de uso que, aunque no tienen valor directamente para los actores principales, son necesarios para que el sistema funcione.

En el contexto del marco de datos personales, se identifican dos **actores principales** que requieren servicios específicos del sistema: usuarios y consumidores de modelos de usuario, y un **actor secundario** al que recurre el sistema para llevar a cabo los casos de uso: fuentes de datos externas.

Los actores principales del marco de datos personales definen el diagrama de casos de uso del sistema que se puede visualizar en la Figura 5-2. Aquí, cada caso de uso se ha representado como un óvalo y un nombre que indica su funcionalidad (parte central) y se identifican los dos actores principales (parte izquierda) y el actor secundario (parte derecha). Además, se han añadido relaciones *include* con una flecha discontinua entre distintos casos de uso que, por ejemplo, cuando va de un caso A a un caso B indica que el caso de uso B puede llevarse a cabo cuando un actor interactúa con el caso de uso A. También se han incluido relaciones de generalización o herencia con una línea continua que simbolizan que un caso de uso especializa uno más general o abstracto al que apunta la flecha.

Figura 5-2. Diagrama de casos de uso del marco de datos personales de referencia.

Del diagrama de casos de uso anterior se puede identificar que los usuarios interactúan con el marco de datos personales para llevar a cabo funcionalidades clásicas de los servicios telemáticos como son: registrarse, darse de baja y acceder al sistema y además, un caso de uso denominado como “Gestionar privacidad” que es característico del sistema. Como se detallará posteriormente en el apartado 5.1.2 Especificación complementaria: requisitos no funcionales, la privacidad supone un requisito no funcional fundamental del sistema pero dado que tiene valor de negocio propio, ésta se modela como un requisito funcional del mismo.

Respecto al actor consumidor, éste interactúa con el marco de datos personales para llevar a cabo dos casos de uso principales: “Recuperar modelos de usuario” y “Solicitar nuevos recursos”. El primero permite que los consumidores obtengan del sistema modelos de usuario que satisfacen sus necesidades de negocio específicas, mientras que la solicitud de nuevos recursos consiste en que éstos puedan indicar qué recursos personales desean que se incluyan en el sistema. Estos dos casos de uso requieren que el usuario autorice a los consumidores para llevar a cabo su ejecución y por lo tanto, que se ejecute el caso de uso “Gestionar privacidad”.

Como se señaló anteriormente, cada uno de los casos de uso principales puede requerir la participación de actores secundarios y la ejecución de otros casos de uso con el objeto de cumplir su propósito. En este sentido, la siguiente figura representa un diagrama de casos de uso extendido del marco de datos personales que incluye, además de los casos de uso principales, los casos de uso secundarios del sistema que están relacionados con actividades específicas del proveedor del sistema y por consiguiente, especifican su comportamiento particular.

Figura 5-3. Diagrama de casos de uso extendido del marco de datos personales de referencia.

El caso de uso “Gestionar privacidad” se ha representado como una especialización del caso de uso “Gestionar usuarios” que permite al sistema administrar los datos personales de los usuarios y centralizar las solicitudes de recuperación e incorporación de recursos que realizan los consumidores de modelos de usuario.

Por otra parte, se han representado otros casos de uso secundarios del sistema, cuyos propósitos son:

- Caso de uso “Recuperar datos personales” permite obtener datos personales de distintas fuentes externas. Este caso de uso se ejecuta cuando un usuario se registra en el sistema utilizando una fuente de datos personales externa o cuando se generan o integran distintos recursos solicitados por el consumidor de modelos de usuario. Los detalles relacionados con este caso de uso se presentarán en el capítulo 6.
- Caso de uso “Generar recursos” permite la creación de nuevos recursos, en concreto, modelos de usuario, dada la solicitud emitida por un consumidor y siempre y cuando se tengan los permisos necesarios por parte del usuario. La ejecución de este caso de uso puede provocar la recuperación de datos personales de distintas fuentes de datos personales externas y la integración de los recursos recuperados o generados. Posteriormente, en el capítulo 7 se detallan los distintos patrones que permiten llevar a cabo este caso de uso al sistema.
- Caso de uso “Integrar recursos” permite registrar los recursos personales de distintas fuentes de datos externas a los que el sistema tiene acceso para recuperar sus datos personales. En el capítulo 8 se describirá una infraestructura para la gestión de identidad que habilita el registro y la vinculación de distintas fuentes de datos externas al sistema con la participación del usuario.

A continuación, se describen los casos de uso principales del sistema especificando los escenarios más representativos del sistema que permiten ilustrar lo que ocurre cuando se instancia un caso de uso particular (Larman, 2001). Como se ha señalado, los casos de uso secundarios que realiza el sistema se describirán en los capítulos 6, 7 y 8.

5.1.1.1 Registro, baja y acceso al sistema

Al igual que otros servicios telemáticos, el marco de datos personales debe ofrecer a sus usuarios funcionalidades relacionadas con el registro, baja y acceso al sistema que le permitan darse de alta, anular su cuenta y autenticarse en el marco de datos personales.

En primer lugar, para que los usuarios puedan acceder a los servicios específicos del marco de datos personales, éstos deben llevar a cabo un proceso de **registro** que les permita darse de alta y obtener sus credenciales de acceso al sistema. Este proceso lo inicializa el usuario a través de una interfaz gráfica del sistema que deberá ofrecer distintos mecanismos y alternativas.

Principalmente, los usuarios podrán darse de alta en el sistema mediante un formulario de registro que deberán completar con sus datos personales, a través de mecanismos de federación de identidades o utilizando su identidad digital de otros proveedores de servicios.

Además, el sistema deberá considerar otras alternativas que puedan surgir con el tiempo y que faciliten el registro a los usuarios.

En el caso opuesto, los usuarios podrán **darse de baja** en el sistema y así, anular su relación con el marco de datos personales. Para ello el usuario deberá acceder con sus credenciales de acceso a una interfaz gráfica que le permite llevar a cabo el proceso de baja, por ejemplo, a través de un formulario o mediante la eliminación manual de los recursos personales vinculados en el sistema. Independientemente del mecanismo utilizado, el proceso de baja de un usuario debe eliminar todos los recursos personales relacionados con éste dentro del marco de datos personales.

Por último, los usuarios podrán **acceder** al marco de datos personales identificándose ante el sistema con sus credenciales o distintas identidades digitales vinculadas en el sistema y esto les permitirá gestionar sus recursos personales y con ello, llevar el caso de uso “Gestionar privacidad” que se detalla a continuación.

5.1.1.2 Gestión de la privacidad

El caso de uso “**Gestionar privacidad**” permite a los usuarios administrar los recursos personales que le refieren para decidir qué datos personales y modelos de usuario se vinculan al sistema, quién los utiliza y para qué. Como resultado de su ejecución, el usuario establece políticas de privacidad que indican los recursos personales que pueden utilizar el resto de actores y el sistema deberá implantar las políticas establecidas.

La inicialización de este caso de uso requiere que los usuarios estén autenticados frente al sistema. Además, para que los usuarios puedan establecer los permisos de acceso y uso de sus recursos personales, éstos deben tener vinculados recursos personales en el sistema. Lo anterior se lleva a cabo de las siguientes formas:

- Los **usuarios pueden aportar datos personales** directamente en el sistema a través de distintos formularios. Así, pueden completar formularios específicos donde trasladen datos personales o también podrán subir archivos estáticos de datos personales obtenidos previamente de distintos proveedores de servicios.
- Los **usuarios pueden vincular fuentes de datos externas** para que el sistema recupere los recursos ofrecidos por éstas. Este proceso de vinculación depende de la fuente de datos considerada y por ello, el sistema indica al usuario el procedimiento a seguir para cada una de ellas. Por ejemplo, existirán fuentes de datos con recursos privados que requieran la autorización del usuario para llevar a cabo la recuperación y otras fuentes de datos públicas en las que el usuario tenga que dar su consentimiento al sistema.
- Dada la vinculación de los recursos anteriores, el sistema podrá llevar a cabo la generación de modelos de usuario que refieran a un usuario concreto si éste da su consentimiento. Por lo tanto, el **usuario deberá permitir la generación de modelos de usuario** y posteriormente, sobre éstos podrá decidir quién los utiliza y para qué.

Dadas las consideraciones anteriores, los usuarios podrán consultar a través de una interfaz gráfica:

- Qué recursos personales están vinculados en el sistema. Esto incluye datos personales aportados por él mismo o procedentes de fuentes externas y modelos de usuario generados por el sistema.
- Disponibles. Todavía puede vincular en el sistema por iniciativa propia. Es decir, puede dar su consentimiento o autorización para que se vinculen en el sistema.
- Solicitados por terceras entidades para su vinculación. Por ejemplo, aquellos que distintos consumidores requieren o el propio sistema para la solicitud de generación de un modelo de usuario.

Sobre esta interfaz, los usuarios podrán otorgar o denegar permisos de acceso y uso de sus recursos personales a distintas entidades, incluido el marco de datos personales.

Además de lo anterior, los usuarios podrán modificar los permisos otorgados, actualizar los recursos personales vinculados (modificando los valores concretos asociados a éstos), visualizar los historiales de uso y acceso a sus recursos, entre otros.

5.1.1.3 Recuperación de modelos de usuario

El caso de uso “**Recuperar modelos de usuario**” permite a los consumidores (entidades externas del sistema) obtener modelos de usuario que satisfacen sus necesidades de negocio específicas. El éxito de este escenario reside en que el sistema cuente con los recursos solicitados o las capacidades necesarias para la generación de los mismos y además, que el usuario haya concedido los permisos de acceso a los consumidores.

Un consumidor inicializa el caso de uso cuando solicita, a través de una interfaz de comunicación provista por el sistema, la recuperación de un modelo de usuario específico. Cuando el sistema recibe dicha solicitud, comprueba las políticas de privacidad establecidas por el usuario al que aplica el modelo de usuario requerido y si tiene los permisos adecuados, entrega el recurso solicitado a través de la interfaz de comunicación. En el caso de que el consumidor no tenga los permisos necesarios, el sistema informa al consumidor y realiza una solicitud de nuevo recurso al usuario, desencadenando el caso de uso “Solicitar nuevos recursos” que se describirá a continuación.

La ejecución del caso de uso de recuperación de modelos de usuario supone para los consumidores de modelos de usuario un procedimiento simple y sencillo, sin embargo requiere la realización de procesos complejos relacionados con la generación de los recursos solicitados por parte del sistema. Por ejemplo, la generación de modelos de usuario implicará que se lleven a cabo la recuperación de datos personales y la ejecución de técnicas de razonamiento, inferencia, agregación, etc. que permitan la creación de nuevo conocimiento del usuario. En el capítulo 7 se detallarán estos procesos.

5.1.1.4 Solicitud de nuevos recursos

Como se ha señalado anteriormente, el caso de uso “**Solicitar nuevos recursos**” puede inicializarlo el sistema cuando un consumidor demanda la recuperación de modelos de usuario y no cuenta con los permisos del usuario adecuados. Además, este caso de uso puede inicializarlo directamente el consumidor de modelos de usuario para obtener nuevos recursos personales de los usuarios con su consentimiento.

Teniendo en cuenta el último caso indicado, un consumidor de modelos de usuario puede solicitar la vinculación de fuentes externas y con ello, de datos personales específicos de los usuarios, y los permisos de acceso asociados para recuperar nuevos modelos de usuario. Adicionalmente, esta solicitud puede tener asociada ofertas o distintos incentivos para estimular la aprobación de los usuarios.

La finalización del caso de uso requiere que el usuario lleve a cabo distintos procesos relacionados con la gestión de la privacidad como, por ejemplo, la vinculación de fuentes de datos o la actualización de sus políticas de privacidad si lo cree conveniente. El sistema deberá informar al consumidor del resultado de su petición y además, incluir los mecanismos necesarios que le permitan notificar al usuario sobre la inicialización del caso de uso de “Gestión de privacidad”. En este sentido, se podrá notificar al usuario en tiempo en real sobre la solicitud realizada o registrar ésta para que el usuario la visualice como una solicitud pendiente cuando acceda al sistema.

5.1.2 Especificación complementaria: requisitos no funcionales

En el apartado anterior se han especificado los requisitos funcionales de un marco de datos personales de referencia mediante la descripción de alto nivel de los casos de uso del sistema. Esto permite establecer los servicios concretos del mismo y su relación con los actores principales y secundarios. Adicionalmente, conviene formalizar una especificación complementaria que capture los requisitos no funcionales del sistema.

Los requisitos no funcionales determinan características y restricciones sobre el sistema en su totalidad, más que rasgos particulares del mismo, aunque influyen sobre el diseño de éste en igual o mayor medida que los requisitos funcionales. De esta forma, los requisitos no funcionales son principios indispensables dentro del diseño de un sistema.

A continuación, se especifican los requisitos no funcionales más relevantes del marco de datos personales: interoperabilidad y privacidad. Además, en el último subapartado se incluyen otros requisitos no funcionales, que aun no siendo objetivos específicos de la tesis doctoral, conviene reflejar.

5.1.2.1 Interoperabilidad

El Instituto de Ingeniería Eléctrica y Electrónica (IEEE, 2010) define la interoperabilidad como la capacidad de un sistema o producto para trabajar con otros sistemas y/o productos sin un esfuerzo especial por parte de sus clientes.

Dada la heterogeneidad existente entre los distintos sistemas como la relacionada con el software, el hardware, los mecanismos de comunicación, la representación de la información, etc. El término interoperabilidad puede tener muchas connotaciones y por ello, generalmente, éste se asocia a distintos niveles y perspectivas que definen las capacidades que han de cumplir los sistemas.

El marco de datos personales interactúa con diferentes sistemas externos de las fuentes de datos personales y los consumidores de modelos de usuario. Así, para su especificación, se deberán tener en consideración distintos niveles de interoperabilidad que faciliten la recuperación e integración de datos personales de distintas fuentes externas, así como la provisión del nuevo conocimiento generado a partir de éstos a terceras entidades. A continuación, se detallan los niveles de interoperabilidad más relevantes para el marco de datos personales, tomando la clasificación de (Callejo, 2009) que distingue 15 niveles de interoperabilidad citados en la literatura.

- **Interoperabilidad semántica:** aborda retos relacionados con el significado de los datos o información intercambiada entre distintos sistemas. Ésta se centra en asegurar que cualquier sistema comprende lo expresado por un dato intercambiado de un modo preciso para su correcta utilización.

El marco de datos personales debe afrontar dos perspectivas relacionadas con la interoperabilidad semántica. Por una parte, el sistema debe incluir técnicas que le permitan comprender los vocabularios y terminologías utilizadas en las fuentes de datos externas que le proveen de datos personales. Además, por otra parte, debe ofrecer soluciones semánticamente interoperables a los consumidores de modelos de usuario para que éstos puedan interpretar correctamente el significado transmitido.

- **Interoperabilidad sintáctica:** se atribuye a aspectos relacionados con el lenguaje, la estructura, formato o lógica utilizada para el intercambio de datos entre dos sistemas.

Al igual que en el caso anterior, el marco de datos personales debe abordar soluciones que permitan alcanzar la interoperabilidad sintáctica en sus interfaces de comunicación con sistemas externos como las fuentes de datos y los consumidores de modelos de usuario.

Ya se ha señalado en la especificación funcional, que los mecanismos de recuperación de los datos personales almacenados en fuentes de datos externas dependen principalmente de los servicios y soluciones incluidos por la fuente considerada, por ello, los esfuerzos relacionados con la interoperabilidad sintáctica del sistema deberán realizarse en la comunicación con los consumidores de modelos de usuario.

- **Interoperabilidad esquemática o estructural:** identifica dificultades relacionadas con la estructura, los esquemas o los modelos de datos utilizados por los distintos sistemas que contribuyen a la falta de interoperabilidad semántica.

Especialmente, el marco de datos personales debe abordar este nivel de interoperabilidad para homogeneizar los recursos personales almacenados en distintas

fuentes y así, facilitar la interoperabilidad semántica de los recursos integrados en el sistema.

- **Interoperabilidad de sistema** o también denominada interoperabilidad técnica: engloba aspectos relacionados con las capacidades de conexión y comunicación de los sistemas que permiten llevar a cabo el intercambio de información o datos.

El marco de datos personales debe incluir servicios y soluciones de interconexión para alcanzar distintas fuentes de datos externas, así como implementar interfaces abiertas que permitan a los consumidores de modelos de usuario obtener y recuperar los recursos requeridos.

En resumen, el marco de datos personales debe considerar soluciones de interoperabilidad dentro de cuatro niveles fundamentales: semántico, sintáctico, estructural y de sistema para permitir y optimizar el intercambio de recursos personales entre los actores participantes. Principalmente, el sistema debe recuperar e integrar datos heterogéneos de distintas fuentes externas, considerando las particularidades de cada una de ellas, y ofrecer una vista homogénea a los consumidores de modelos de usuario que requieren nuevo conocimiento de los usuarios.

Aunque pueden especificarse como un requisito no funcional independiente, la **utilización de normas, estándares, vocabularios comunes y la promoción y uso de software de código abierto** garantizan y fomentan la interoperabilidad de los sistemas, especialmente en los niveles descritos, y por ello, el diseño y la implementación del marco de datos personales debe considerar dichas soluciones.

Alcanzar los niveles de interoperabilidad identificados determinará la generación de valor para los actores participantes. En capítulo 6 se describen las contribuciones y validaciones relacionadas con la interoperabilidad de los recursos personales del marco de datos personales.

5.1.2.2 Privacidad

Tal y como se señaló en el capítulo 3 Estado del arte (3.5.1 Privacidad desde el diseño), la privacidad debe entenderse como característica global del sistema que considere todas las áreas de un servicio o producto, incluyendo dentro de éstas las prácticas de negocio, el diseño de las infraestructuras físicas y la aplicación de tecnologías.

Desde una perspectiva de negocio, el marco de datos personales se ha descrito bajo un enfoque de plataforma multilateral (3.6.2 Modelos de negocio en el contexto de servicios telemáticos) y así, se establece en torno a distintas propuestas de valor que promueven prácticas de negocio responsables en relación al uso y gestión de datos personales de los usuarios. Estas propuestas han sido trasladadas a la especificación funcional del sistema (5.1.1 Descripción funcional de alto nivel mediante casos de uso) y en concreto, se ha definido que el sistema debe permitir a los usuarios gestionar la privacidad de sus recursos personales y por otra parte, que debe implantar las políticas de privacidad establecidas por éstos para evitar que entidades no autorizadas hagan uso de los recursos personales de los usuarios.

De las consideraciones anteriores se desprende que la privacidad está siendo considerada desde el inicio del desarrollo del marco de datos personales, tanto en la definición de su modelo de negocio, como en su especificación funcional y por consiguiente, siguiendo la filosofía descrita por la doctrina de privacidad desde el diseño. No obstante, dicha doctrina gira en torno a siete principios fundamentales que deben entenderse como siete reglas de domino para el marco de datos personales y que deberán considerarse a lo largo de todo el ciclo de vida del sistema.

En el capítulo 8 se presenta la infraestructura de privacidad establecida para el marco de datos personales y que tendrá en consideración los principios de privacidad desde el diseño que hasta el momento no se han destacado.

5.1.2.3 Otros requisitos no funcionales

Además de los requisitos no funcionales descritos anteriormente, el marco de datos personales debe considerar otros relacionados con la escalabilidad, automatización de los procesos, reusabilidad, ubicuidad, fiabilidad, rendimiento, usabilidad y accesibilidad, aunque éstos no son objeto directo de la tesis doctoral y por consiguiente, no se detallan.

5.2 Arquitectura de referencia

Una vez se ha formalizado la especificación del marco de datos personales, se está en disposición de establecer el diseño de una arquitectura de referencia que podrá ser particularizada en diversos ámbitos de aplicación y mediante la utilización de diferentes tecnologías.

La Figura 5-4 representa dicha arquitectura detallando los módulos principales que la conforman, las relaciones entre éstos y las distintas interacciones con los actores externos: fuentes de datos personales (distinguiendo entre públicas y privadas), consumidores de modelos de usuario y los propios usuarios del sistema.

Figura 5-4. Arquitectura de referencia del marco de datos personales (San Miguel, et al., 2015a).

En la figura se distinguen seis módulos que se corresponden con los requisitos funcionales y no funcionales establecidos en la sección anterior. A continuación, se describe cada uno de los módulos en detalle, señalando sus características más relevantes y los requisitos que llevan a cabo.

5.2.1 Gestor de usuarios

El módulo Gestor de usuarios es un módulo vertical que permite al usuario interactuar con el marco de datos personales para inicializar los casos de uso en los que actúa como actor principal. Así, a través de una interfaz gráfica gestionada por el módulo, los usuarios pueden registrarse, darse de baja, acceder al sistema y gestionar su privacidad. La gestión de privacidad y conforme a lo descrito en 5.1.1.2, implica que, si así lo desea el usuario, éste aporte datos personales, que consienta o autorice la recuperación de recursos al sistema y que permita la generación de modelos de usuario.

Dadas las consideraciones anteriores, el módulo Gestor de usuarios incluye y administra distintas interfaces y formularios para que el usuario lleve a cabo tareas relacionadas con el abastecimiento de recursos en el sistema y por otra parte, con la gestión de sus identidades digitales y la privacidad de sus recursos. Para la ejecución de éstas, el módulo se comunica con los módulos de la arquitectura Recolector y Registro a los que reenvía las órdenes indicadas por el usuario.

Posteriormente, en el capítulo 6 se detallarán distintos mecanismos que permiten llevar a cabo la recuperación de recursos al sistema a través del módulo Recolector y cómo éste se apoya en el módulo Gestor de usuarios, bien para recibir datos directamente del usuario, o bien para que éste consienta y autorice la recuperación de recursos custodiados por fuentes externas.

Además, en el capítulo 8 se establece una infraestructura para la gestión de la identidad y privacidad centrada en el usuario que se sustenta en el módulo Gestor de usuarios para establecer las preferencias del usuario en el sistema.

Por último, cabe señalar que, dado que el módulo Gestor de usuarios se encarga de la comunicación con el usuario, este módulo debe considerar principios de diseño universal e incorporar interfaces gráficas que faciliten la interacción del usuario de forma intuitiva e independientemente de sus características particulares. Aunque lo anterior no se detalla explícitamente en los capítulos de la tesis, sí se han realizado distintos trabajos relacionados como los señalados en el capítulo 9 del proyecto Segur@ y el proyecto Buscamedia o los incluidos en los artículos (Yelmo, et al., 2006), (Yelmo, et al., 2007a), (Yelmo, et al., 2007b), (Martín, et al., 2009), (San Miguel, et al., 2012) y (Yelmo, et al., 2010).

5.2.2 Recolector

El módulo Recolector es el encargado de obtener y recuperar datos personales de distintas fuentes externas: el propio usuario, fuentes de datos públicas o fuentes de datos privadas. Para ello, el módulo se descompone en distintos componentes específicos que permiten la recuperación de cada una de las fuentes de datos consideradas y en aquellos que llevan a cabo la activación y orquestación de los anteriores.

En el capítulo 6 se analizarán en profundidad los distintos componentes del módulo Recolector, así como las relaciones que establece con el resto de módulos del sistema. Principalmente, este módulo está formado por componentes específicos que permiten la recuperación de recursos a través de cuatro mecanismos principales (3.3 Mecanismos de recuperación de datos personales almacenados en fuentes de datos externas): programas araña, ficheros estáticos, basados en peticiones y basados en suscripciones, u otros que pueden detectarse como, por ejemplo, aquellos asociados con la recuperación de datos aportados por los usuarios.

Por otra parte, respecto a la interacción con el resto de módulos, el módulo Recolector se comunica con el módulo Registro para comprobar los permisos y políticas de privacidad establecidas por el usuario y conforme a esto, lleva a cabo la activación de sus componentes para realizar la recuperación de recursos. Además, los recursos recuperados por el módulo pueden redirigirse al módulo Almacén de recursos o Generador para llevar a cabo la generación de modelos de usuario tal y como se detallará en el capítulo 7.

5.2.3 Registro

El módulo Registro es el encargado de ejecutar las peticiones realizadas por los usuarios a través del módulo Gestor de usuarios, así como de establecer las comunicaciones con los proveedores de servicios externos (fuentes de datos y consumidores de modelos) para llevarlas a cabo.

Por una parte, el módulo Registro actúa como un proveedor de identidad que permite a los usuarios registrarse en el sistema y vincular sus identidades digitales de otros proveedores.

Además, actúa como gestor de las políticas de privacidad establecidas por el usuario llevándolas a cabo. Para lo anterior, se apoya en el módulo Almacén de recursos donde dirige los datos personales aportados por los usuarios, la información y permisos asociados con otras identidades digitales y las políticas de privacidad establecidas por el usuario. De esta forma, cuando recibe una petición de un consumidor a través del módulo Recuperador realiza los procedimientos adecuados para implantar las políticas del usuario. En el capítulo 8 se detallará esta funcionalidad del módulo Registro destacando los componentes que lo conforman. Cabe destacar que deberá incluir módulos que permitan la traducción de políticas de privacidad de usuario y aquellos que permitan su evaluación.

Por otra parte, este módulo interactúa con el módulo Recolector y Generador para obtener los recursos necesarios en la generación de modelos de usuario, tal y como se detallará en los capítulos 6 y 7.

En último lugar, este módulo es también el encargado de recibir las peticiones de distintos consumidores que requieren modelos de usuario no considerados en el sistema y de comunicárselo al usuario a través de un sistema de notificaciones, como se señalará en el capítulo 8.

5.2.4 Generador

El módulo Generador permite obtener nuevo conocimiento de los usuarios a partir de los datos personales recuperados en el sistema para producir modelos de usuario que se adaptan a las necesidades de negocio particulares de los consumidores.

Para lo anterior, el módulo Generador está formado por componentes específicos que permiten la producción de modelos de usuario conforme a las preferencias del usuario. Tal y como se detallará en el capítulo 7, el establecimiento de estos componentes depende principalmente de dónde se encuentren los recursos necesarios para la generación y de las características de los modelos de usuario solicitados por los consumidores. De esta forma, podrán considerarse componentes específicos con enfoques basados en el procesamiento complejo de eventos o técnicas de inferencia, entre otros.

Dadas las consideraciones anteriores, el módulo Generador se comunica con el módulo Registro para obtener los permisos del usuario respecto a la generación de modelos y/o los necesarios para iniciar la recuperación de recursos. Además, se comunica con el módulo Recolector y/o el Almacén de recursos para obtener los datos necesarios en la generación de modelos. Por último, los modelos de usuario podrán reenviarse al módulo Almacén de recursos o bien podrán entregarse a los consumidores de modelos a través del módulo Recuperador.

Como se ha mencionado, los detalles concretos sobre la interacción del módulo Generador con el resto de módulo, así como los componentes específicos que puede considerar se describirán en el capítulo 7.

5.2.5 Almacén de recursos personales

El módulo Almacén de recursos personales recopila información relacionada con los datos personales, modelos de usuario y las identidades digitales de los usuarios. De esta forma, puede custodiar datos personales recuperados de otras fuentes y/o modelos de usuario generados en el sistema. Además, almacena información necesaria para la gestión de identidades y privacidad como los permisos de uso y acceso otorgados por los usuarios, las políticas de privacidad que éstos establecen, sus identidades digitales creadas en otros proveedores, etc.

Este módulo incluye una interfaz de comunicación donde distintos módulos de la arquitectura redirigen u obtienen recursos necesarios para llevar a cabo sus funcionalidades. Por una parte, tal y como se describirá en el capítulo 6, el módulo Recolector puede redirigir los recursos recuperados al módulo Almacén de recursos para su almacenamiento. Además, como se detallará en el capítulo 7, el módulo Generador puede obtener datos almacenados en este módulo para llevar a cabo la generación y redirigir los modelos de usuario producidos para su almacenamiento. Por otra parte, el módulo Registro se comunica con el Almacén de recursos para guardar información relacionada con los permisos de autorización y políticas de privacidad de los usuarios tal y como se ha señalado y como se detallará en el capítulo 8.

Principalmente, se deberán considerar en este módulo soluciones interoperables que favorezcan la generación de modelos de usuario y la combinación de las identidades digitales de los usuarios.

5.2.6 Recuperador

El módulo Recuperador es el encargado de implementar los mecanismos de comunicación necesarios para que los consumidores de modelos puedan solicitar y recuperar recursos con el consentimiento y la autorización del usuario.

De forma análoga al módulo Recolector, el módulo Recuperador estará formado por diferentes componentes específicos que permiten la recuperación de los recursos a través de distintos mecanismos. Principalmente, se considerarán mecanismos basados en peticiones, donde un consumidor realiza peticiones puntuales para la recuperación de modelos de usuario, o mecanismos basados en suscripciones que permitan la obtención y generación de modelos a medida que se generan.

El módulo Recuperador se comunica con el módulo Registro para comprobar los permisos de acceso establecidos por el usuario y si los permisos indican que el consumidor puede acceder al modelo de usuario solicitado se responde a los consumidores con el recurso solicitado o se informa de la suscripción a los mismos. Lo anterior requiere que el módulo Recuperador se comunique con el módulo Almacén de recursos o el módulo Generador para obtener los modelos de usuario solicitados.

En el capítulo 7 se presentan los detalles relacionados con el módulo Recuperador y las interacciones que puede establecer con el resto de módulos de la arquitectura para responder a los consumidores con los modelos de usuario solicitados. Asimismo, en el capítulo 8 se

describe la participación de este módulo para llevar a cabo el establecimiento de políticas de privacidad de los usuarios que permitirán o denegarán el acceso a modelos de usuario solicitados.

5.3 Resumen y contribuciones del capítulo

En este capítulo se ha establecido la especificación y el diseño de un marco de datos personales de referencia considerando distintos enfoques y sin hacer alusión a ninguna tecnología concreta. De esta forma, dependiendo del ámbito de aplicación y del propósito particular del marco de datos personales, la implementación final de éste podrá llevarse a cabo considerando parte de los enfoques de diseño señalados y mediante distintas soluciones tecnológicas. Por ejemplo, tal y como se presentará más adelante, podrán existir soluciones basadas en el almacenamiento de datos personales en un repositorio central u otras que únicamente consideren recursos custodiados por fuentes externas.

Dadas las consideraciones anteriores, la contribución original incluida en este capítulo es la especificación y diseño de una arquitectura genérica de referencia que se sustenta en el modelo de negocio presentado en el capítulo anterior y que permite la recuperación de recursos custodiados por distintas entidades (fuentes de datos) para generar modelos de usuario enriquecidos que cubren las necesidades específicas de terceras entidades, considerando la participación del usuario y el control sobre sus recursos personales (datos y modelos de usuario). Todo lo anterior conforme a dos principios descritos: interoperabilidad y privacidad desde el diseño.

Esta contribución se ha llevado a cabo principalmente en el proyecto POSDATA del Center for Open Middleware donde se han abstraído los resultados obtenidos en otros proyectos de investigación como el proyecto Segur@, el proyecto Buscamedia (Buscamedia, 2015) y el proyecto MOET en los que ha participado la autora y donde se había abordado distintos enfoques de diseño.

La contribución incluida en este capítulo ha propiciado la publicación de distintos artículos entre los que cabe mencionar:

- *A Personal Data Framework for Distributed User Modelling in New Service Ecosystems* (San Miguel, et al., 2015e) artículo confirmado para su publicación en la revista IEEE América Latina indexada en JCR SCI. Aquí, se incluye una descripción de la arquitectura de referencia presentada y una particularización de ésta basada en la utilización de tecnologías semánticas.
- *Extracting Knowledge through Friendsourcing: Application to new Financial Services* (San Miguel, et al., 2015c) artículo presentado en el VII Congreso Iberoamericano de Telemática (CITA2015) y seleccionado dentro de los cinco mejores del congreso. El artículo presenta la arquitectura de referencia y su implementación para la promoción de nuevos servicios financieros.

- *A Personal Data Framework for Exchanging Knowledge about Users in New Financial Services* (San Miguel, et al., 2015a) artículo presentado en el *1st International Workshop on Personalization & Recommender Systems in Financial Services* (FINREC). Incluye una descripción de la arquitectura presentada y un análisis de su utilidad para ofrecer servicios financieros basados en la personalización y recomendación de contenidos.
- *Personal Data Usage and Management* (San Miguel, et al., 2014a) artículo presentado en el *Doctoral Consortium on Enterprise Information Systems* (DCEIS 2014) donde se detalla la metodología utilizada para el establecimiento de la arquitectura descrita en este capítulo.

Además de lo anterior, los artículos (Yelmo, et al., 2011a) y (Yelmo, et al., 2010) engloban los trabajos realizados dentro del proyecto Buscamedia en el que se describe un marco genérico para ofrecer modelos de usuario a un servicio de búsquedas personalizadas que se basa en la recolección de recursos de distintas fuentes de datos públicas, la generación de modelos de usuario y su aplicación concreta en el servicios de búsquedas. Estos trabajos han servido de base para el establecimiento de la arquitectura de referencia presentada en este capítulo. Cabe señalar que (Yelmo, et al., 2011a) se presentó en el VI Congreso Ibero-americano de Telemática (CITA2011) y por su parte, (Yelmo, et al., 2010) en el *Second International ICST Conference, UCMedia 2010*, donde fue seleccionado entre los mejores del congreso y así, su versión extendida y revisada se incluyó como capítulo de libro.

Finalmente, las publicaciones (Yelmo, et al. 2006), (Yelmo, et al. 2007a), (Yelmo, et al. 2007b) y (Martín, et al., 2009) en las que ha participado la autora de esta tesis incluyen distintos estudios y análisis de requisitos no funcionales relacionados con la accesibilidad y usabilidad que ha permitido establecer el enfoque de diseño para el módulo Gestor de usuarios. Entre los méritos relacionados con estas publicaciones destaca el *John M Slatin Award for Best Communication Paper* al artículo (Martín, et al., 2009).

5.3.1 Enfoque de validación global

El marco de datos personales es una solución ambiciosa que cubre distintos retos relacionados con la recolección, generación, recuperación, gestión de identidad y privacidad de datos personales y modelos de usuario. En el contexto de la tesis doctoral, se han validado los principales módulos de la arquitectura de referencia, así como muchos de los componentes específicos de éstos a través de la implementación de pruebas de concepto y prototipos en distintos escenarios que cubren globalmente los requisitos establecidos. En el resto de capítulos se detallarán las contribuciones y validaciones concretas para cada uno de los módulos de la arquitectura establecida.

En primer lugar, en el **capítulo 6** se profundizará en los módulos relacionados con la recuperación de recursos personales. Esto incluye principalmente el módulo Recolector de la arquitectura de referencia presentada y los módulos Gestor de usuarios y Registro que permiten materializar la participación del usuario en el proceso.

Posteriormente, el **capítulo 7** incluye los trabajos relacionados con la generación y recuperación de modelos de usuario y por lo tanto, se presentarán los detalles relacionados con los módulos Generador, Recuperador y Almacén de recursos.

Por último, el **capítulo 8** incluye las contribuciones relacionadas con la gestión de la identidad y la gestión de privacidad en el sistema que se establece en torno a los módulos Gestor de usuarios, Registro, Almacén de recursos y Recuperador.

Capítulo 6.

Recuperación de datos en el marco de datos personales

A diferencia del proceso de modelado de usuario tradicional (3.2 Proceso de modelado de usuario), y tal y como se describió en el capítulo 4 (4.2.4 Cadena de valor), el marco de datos personales que se propone en esta investigación se sustenta en la distribución de dicho proceso y considera la participación de otras entidades (proveedores de servicios y usuarios) para llevarlo a cabo. Esto permite, por una parte, obtener modelos de usuario que difícilmente pueden generarse de forma unilateral y por otra, abstraer del proceso de modelado de usuario a entidades externas que requieren conocimiento sobre los usuarios, todo ello con el consentimiento expreso de los usuarios y por consiguiente, siendo respetuosos con las preferencias e intereses de privacidad de éstos.

En esta tesis doctoral se define el **proceso de modelado de usuario distribuido** como la realización de dicho proceso aprovechando los recursos que distintas entidades (fuentes de datos) ofrecen para generar modelos de usuario enriquecidos que cubren las necesidades específicas de proveedores de servicios (consumidores), considerando la participación del usuario y el control sobre sus recursos personales (datos y modelos de usuario).

Lo anterior implica abordar distintos retos para cada una de las etapas de proceso de modelado de usuario tradicional (recolección, generación y aplicación) y ampliar los métodos habituales para involucrar la participación de entidades externas. En este sentido, el presente capítulo aborda los relacionados con la primera etapa del proceso (recolección de datos) y posteriormente, en el Capítulo 7 se tratarán los relacionados con las dos últimas etapas (generación y aplicación de modelos de usuario).

En el proceso de modelado de usuario distribuido, la etapa de recolección se transforma en una nueva etapa que ha sido denominada como **recuperación de datos**, cuyo objeto es obtener los recursos necesarios, almacenados en y custodiados por entidades externas, para la generación de modelos de usuario.

Tomando como base la arquitectura de referencia descrita en el capítulo anterior, la siguiente figura señala las distintas etapas del proceso de modelado de usuario distribuido (parte derecha) y las entidades y módulos que permiten llevarlo a cabo (parte izquierda). Aquí, se observa como la etapa de recuperación abarca lo relacionado con las fuentes de datos externas y la participación del módulo Recolector. Además, la etapa de generación se lleva a cabo a través de tres módulos principales: Registro, Almacén de recursos personales y Generador. Por último, la etapa de aplicación identifica la participación del módulo Recuperador y los distintos consumidores de modelos de usuario. De forma transversal a cada una de las etapas, el usuario participa en el proceso a través del módulo Gestor de usuarios.

Figura 6-1. Proceso de modelado de usuario distribuido en el marco de datos personales.

En las siguientes secciones se presentan las contribuciones y validaciones relacionadas con la etapa de recuperación de datos que se han llevado a cabo en la presente investigación. En primer lugar, las secciones 6.1 y 6.2 incluyen respectivamente la identificación y caracterización de las fuentes de datos externas con potencial de abastecer al sistema. Seguidamente, la sección 6.3 presenta una visión global de la etapa de recuperación de recursos en el marco de datos personales, considerando las distintas fuentes de datos identificadas, sus relaciones e interacciones con los módulos y componentes de la arquitectura del sistema involucrados. La sección 6.4 engloba las validaciones realizadas para la etapa de recuperación que se ha llevado a través de distintos escenarios planteados en esta tesis doctoral. Para concluir, la sección 6.5 resume las contribuciones y trabajos incluidos en este capítulo.

6.1 Identificación de fuentes de datos externas

Tal y como se ha señalado anteriormente, y conforme a lo expuesto en 5.1 Especificación del marco de datos personales, la primera etapa del proceso de modelado de usuario distribuido se basa en la recuperación de recursos disponibles en entidades externas al marco de datos

personales con la participación y el consentimiento expreso del usuario. De esta forma, el usuario es el que determina qué datos se recuperan y para qué y en todo momento es consciente del hecho de que está proporcionando o facilitando el acceso a datos personales que le refieren y de las implicaciones que esto supone respecto a su privacidad.

Dadas las consideraciones anteriores, los usuarios pueden participar de forma activa en la etapa de recuperación proporcionando datos directamente en el marco de datos personales, o bien dar su consentimiento y los permisos necesarios para que el sistema lleve a cabo la recuperación de los datos custodiados por distintos proveedores de servicios. Así, se identifican como **fuentes de datos externas** al marco de datos personales: al **propio usuario** y a cada uno de los **proveedores de servicios** a los que el sistema tiene acceso porque así lo ha establecido (consentido o autorizado) el usuario.

Como se indicó en 3.3 Mecanismos de recuperación de datos personales almacenados en fuentes de datos externas, un proveedor de servicios puede clasificarse en función del acceso a sus datos personales como fuente de datos pública y/o privada y esto repercute en los mecanismos a utilizar para la recuperación de sus recursos. Además, los proveedores de servicios concretos que pueden actuar como fuentes de datos externas del marco de datos personales se han clasificado en función de la utilidad de los recursos que custodian como: **fuentes de datos personales y fuentes de datos para el enriquecimiento del proceso de modelado de usuario**. Las primeras son proveedores de servicios que poseen datos personales directamente relacionados con el usuario y que por consiguiente, pueden utilizarse para modelar al usuario concreto al que refieren dichos datos. Por su parte, las fuentes de datos para el enriquecimiento del modelado de usuario proveen distintos recursos como, por ejemplo, datos estadísticos sobre un grupo de usuarios o relaciones entre distintos conceptos, que facilitan y mejoran la producción de modelos de usuario.

Para ilustrar lo anterior, una fuente de datos personales puede custodiar el código postal de un usuario concreto o la localización en la que se encuentra éste en un determinado momento, mientras que una fuente de datos con recursos para el enriquecimiento del proceso de modelado de usuario puede aportar información relacionada con el nivel socio-cultural medio asociado a un código postal o la celebración de eventos deportivos asociados con una determinada localización. Dichos datos pueden relacionarse para deducir o inferir características de los usuarios.

La Tabla 6-1 resume las fuentes de datos externas identificadas señalando la utilidad de cada una de ellas en el proceso de modelado de usuario conforme a lo expuesto anteriormente. Es importante señalar que, tal y como se describirá en 6.2.3 Otras fuentes de datos y recursos disponibles, se han identificado dos recursos disponibles (información de comunidad e información de contexto) como fuentes de datos externas dada su relevancia pero que éstos estarán custodiados por distintos proveedores de servicios.

Tabla 6-1. Fuentes de datos externas identificadas y utilidad en el proceso de modelado de usuario.

FUENTE DE DATOS EXTERNA	UTILIDAD
Usuario	Aporta datos que le refieren Consiente y/o autoriza la recuperación en el resto
Redes sociales	Custodia datos del usuario
Servicios de mensajería	Custodia datos del usuario
Proveedores asociados a dispositivos móviles	Custodia datos del usuario
Brókeres de datos	Custodia datos del usuario
Proveedores de servicios de Internet de las cosas	Custodia datos del usuario
Proveedores específicos privados	Custodia datos del usuario
Proveedores específicos públicos	Custodia datos del usuario
Plataforma de datos abiertos	Custodia recursos relevantes para el modelado
Plataforma de datos enlazados	Custodia recursos relevantes para el modelado
Información de comunidad	Recurso relevante para el modelado disponible en distintos proveedores
Información de contexto	Recurso relevante para el modelado disponible en distintos proveedores

A continuación, en la sección 6.2, se caracterizan las fuentes de datos externas con potencial de abastecer al marco de datos personales de los recursos necesarios: el usuario, proveedores de servicios y otras fuentes de datos y recursos disponibles que pueden utilizarse para el enriquecimiento del proceso de modelado de usuario. En el apartado 6.2.4 se incluye un resumen de las características halladas para cada fuente.

6.2 Caracterización de las fuentes de datos externas

Cada una de las fuentes de datos identificadas que tienen potencial de abastecer de los recursos necesarios al marco de datos personales para llevar a cabo el proceso de modelado de usuario distribuido será caracterizada conforme a los siguientes criterios:

- **Recursos custodiados.** En el estado del arte (3.1.1 Datos personales) se puso de manifiesto que los datos personales tienen una complejidad intrínseca derivada de su heterogeneidad y de las múltiples perspectivas desde las que pueden ser evaluados y clasificados. Para la caracterización de las fuentes de datos externas se tomará como referencia las diez categorías generales expuestas en el estado del arte (FIDIS, 2005):
 - información de identificación,
 - características biológicas,

- localización,
- activos y transacciones,
- ciudadanía y justicia,
- preferencias,
- competencias,
- funciones, posiciones y roles,
- características sociales, y
- características psicológicas.

Así, para cada fuente de datos externa deberá indicarse qué categorías de datos personales ofrecen y asimismo, se señalarán, cuando sea relevante su distinción, aspectos relacionados con la sensibilidad, temporalidad o el origen de creación de dichos datos (3.1.1 Datos personales).

- **Tipo de acceso a los recursos custodiados.** Cada una de las fuentes de datos considerada se identificará conforme al acceso a sus recursos personales como: pública y/o privada (3.3 Mecanismos de recuperación de datos personales almacenados en fuentes de datos externas). Dado que dicha clasificación depende del recurso considerado, se deberá señalar para cada fuente de datos si ofrece datos con acceso público, privado o ambos.
- **Mecanismos para la recuperación de los recursos custodiados.** Tal y como se señaló en el estado del arte (3.3 Mecanismos de recuperación de datos personales almacenados en fuentes de datos externas), la recuperación de los recursos custodiados por una fuente de datos puede llevarse a cabo a través de cuatro mecanismos principales: programas araña, ficheros estáticos, peticiones y suscripciones. La aplicación de cada uno de estos mecanismos depende del tipo de acceso a dichos recursos (público y privado) y de los servicios específicos incluidos por la fuente de datos considerada que posibilitan o habilitan dichos mecanismos. Para cada una de las fuentes de datos identificadas deberán indicarse los mecanismos de recuperación asociados, así como los servicios que incluyen las distintas fuentes de datos para habilitar la recuperación.

Cabe señalar que la recuperación de datos privados requiere e implica que una entidad competente (el usuario o la propia fuente de datos) dé su autorización para llevar a cabo su recuperación, mientras que para los datos públicos puede llevarse a cabo libremente y sin ningún permiso especial. En el contexto de esta investigación, se considera que el usuario debe dar su consentimiento explícito o autorización para la aplicación de cualquiera de los mecanismos identificados, independientemente de si estos datos son públicos o privados.

En los siguientes apartados se describen las fuentes de datos identificadas conforme a los criterios y consideraciones establecidos.

6.2.1 El usuario como fuente de datos

Los usuarios del marco de datos personales pueden generar y revelar datos y conocimiento sobre sí mismos y por lo tanto, éstos pueden actuar como una fuente de datos externa con potencial de abastecer al sistema.

Dentro de dicha perspectiva, el marco de datos personales es el responsable de implementar los mecanismos que permitan a los usuarios aportar los datos personales que le refieren. En este sentido, se establece que los usuarios podrán crear o generar datos activamente y directamente dentro del sistema, o bien podrán recuperar datos custodiados por otros proveedores de servicios para importarlos o incluirlos en el mismo y así, reducir el número de tareas manuales a realizar.

Al igual que en el proceso de modelado de usuario tradicional (ver 3.2.), en el marco de datos personales, los usuarios pueden aportar los datos requeridos para el modelado de usuario rellenando distintos formularios que permiten obtener información específica de éstos, por ejemplo, información relacionada con el registro (como nombre o dirección de correo electrónico) u otra más concreta, como preferencias musicales, datos laborales, etc. El establecimiento de estos formularios depende de las necesidades concretas de los consumidores de modelos de usuario y con ello, de los datos requeridos para la producción de los modelos solicitados.

Además de los formularios, el marco de datos personales permite que los usuarios importen datos de otros proveedores de servicios a través de ficheros estáticos obtenidos en dichos proveedores. Lo anterior implica que los usuarios realicen distintas tareas manuales en el proveedor origen, que dependen de este último, para obtener los datos requeridos. Por ejemplo, Google, Facebook o Twitter permiten a sus usuarios dentro de las configuraciones u opciones avanzadas de sus servicios exportar y descargar datos personales en distintos archivos. Además, la solicitud de acceso a ficheros de datos personales, así como la rectificación, cancelación y oposición de los datos de éstos, también se puede realizar en otros proveedores de servicios a través de diferentes canales bajo una acreditación del solicitante (Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal).

Parece obvio pensar que la información aportada directamente a través de los usuarios será más precisa y veraz que la información que se puede obtener de otras fuentes externas o inferir a partir de los datos recolectados. No obstante, hay que considerar que los mecanismos descritos anteriormente requieren que los usuarios lleven a cabo dichas tareas manuales y por lo tanto, que tengan que dedicar cierto tiempo a completarlas. Los usuarios no siempre estarán dispuestos a colaborar rellenando formularios o descargando archivos en distintos proveedores de servicios y con ello, la utilización de estos mecanismos puede llegar a sobrecargarlos, haciendo que pierdan su interés o que las molestias ocasionadas no compensen los beneficios obtenidos, lo cual puede traducirse a que introduzcan datos incorrectos o incluso a que no lleven a cabo dichas tareas.

Para evitar lo anterior, el marco de datos personales debe alcanzar un equilibrio entre el número de tareas manuales a realizar por parte del usuario y la compensación obtenida por éstos. En definitiva, el sistema debe evitar sobrecargar a sus usuarios con tareas manuales y en este sentido, el marco de datos personales incluye soluciones que permiten delegar a los usuarios la recuperación de datos personales de distintos proveedores de servicios. De esta forma, se reduce el tiempo que los usuarios deben invertir para aportar los datos requeridos.

6.2.2 Proveedores de servicios como fuentes de datos personales

Como se ha señalado anteriormente, los usuarios pueden establecer que el marco de datos personales lleve a cabo la recuperación de los datos necesarios para el proceso de modelado de usuario y así, reducir el número de tareas manuales que deben realizar. Para ello, los usuarios deben dar el consentimiento y/o los permisos necesarios que permitan al marco de datos personales acceder a los datos custodiados en distintos proveedores de servicios.

De este modo, los proveedores de servicios se establecen como fuentes de datos personales del sistema cuando la recuperación de sus recursos la realiza directamente el sistema con el consentimiento y/o autorización del usuario.

A continuación, se describen los proveedores de servicios más significativos del panorama actual que tienen potencial de actuar como fuentes de datos externas del marco de datos personales. Cabe destacar que cada uno de los proveedores de servicios considerados en este apartado incluye **información de identificación** de los usuarios, puesto que ésta es necesaria para gestionar la identidad de los mismos dentro de sus servicios y aún más, para autorizar el acceso y la recuperación de los recursos custodiados que refieren a un usuario concreto.

6.2.2.1 Redes sociales

Las redes sociales juegan un papel fundamental dentro del contexto de esta tesis doctoral, puesto que muchas de las validaciones realizadas se basan en éstas. Anteriormente, en 4.3.1 Modelo de negocio de intermediación de publicidad en redes sociales, se describió el origen, las características principales y la relevancia de las redes sociales en el contexto actual y en particular, en los modelos de negocio basados en publicidad. En este apartado se caracterizan conforme a los criterios establecidos en la recuperación de datos personales.

Las redes sociales son servicios telemáticos que permiten a los usuarios participar en comunidades virtuales e interactuar con sus amigos u otras personas con las que comparten intereses comunes (Real Academia de Ingeniería, 2015). Ejemplo de estos servicios son: Facebook, Twitter, Google+⁵⁸, Tuenti⁵⁹ u otras redes especializadas como Instagram⁶⁰ o LinkedIn⁶¹.

Dada la esencia de estos servicios, basados en la generación de datos personales por parte de los usuarios, los recursos personales que custodian dichos servicios proporcionan un rápido

⁵⁸ Google+, <https://plus.google.com/>

⁵⁹ Tuenti, <http://tuenti.com/>

⁶⁰ Instagram, <http://instagram.com/>

⁶¹ LinkedIn, <http://linkedin.com/>

conocimiento de los usuarios y pueden incluir cualquiera de las categorías establecidas, puesto que esto depende de la actitud de los usuarios respecto a la compartición de su información. Principalmente, las redes sociales engloban características sociales y psicológicas, preferencias y roles dentro de una comunidad de un usuario, aunque, por ejemplo, pueden incluir características biológicas si se trata de una red social de temática centrada en salud y/o medicina, información sobre competencias si el usuario comparte un título obtenido, etc.

La mayoría de las redes sociales incluyen perfiles de usuario básicos de acceso público (normalmente, formados por el nombre, apellidos, edad, foto de identificación o una breve descripción textual) y perfiles más completos habitualmente de acceso privado (que pueden incluir listas de amigos, grupos de interés, gustos de los usuarios, eventos, pensamientos, contenidos multimedia, etc., dependiendo del tipo de red social).

En términos generales, el acceso a los datos personales de un usuario por parte de terceras entidades depende de éste y de la configuración y permisos de acceso que seleccione sobre su perfil de usuario (público o privado). Se da la casuística de existir redes sociales con la mayoría de su contenido con acceso público, como por ejemplo Twitter o Instagram, y otras con perfiles más cerrados, como Facebook o Tuenti.

Como puede deducirse de lo anterior, los mecanismos para la recuperación de los datos personales custodiados por los servicios de redes sociales dependen de los datos considerados y en concreto, de si éstos son públicos o privados. De esta forma, se podrán utilizar programas araña para recuperar perfiles públicos, de igual manera que realizan la mayoría de los motores de búsqueda. Además, en cuanto a los perfiles privados, la mayoría de redes sociales permiten a sus usuarios descargar ficheros estáticos con los datos personales generados dentro de sus servicios e incluyen un conjunto de APIs, como, por ejemplo, las señaladas en 3.3.3.1 Recuperación de datos mediante interfaces de programación de aplicaciones, que permiten a terceras entidades acceder a los datos personales privados de un usuario con el consentimiento de éste.

Tomando como referencia los servicios ofrecidos por Facebook y Twitter, las APIs de las redes sociales implementan mecanismos basados en peticiones y suscripciones. De esta forma, una entidad externa puede solicitar la recuperación de un dato privado de un usuario a través del Graph API de Facebook o el REST API de Twitter, o bien puede utilizar la funcionalidad Real Time Updates del Graph API o las Streaming APIs de Twitter para que la red social notifique lo que ocurre en tiempo real dentro de su comunidad o a un usuario particular.

6.2.2.2 Servicios de mensajería instantánea

Los servicios de mensajería instantánea permiten el intercambio de mensajes de texto u otros contenidos (por ejemplo, fotografías, audios o vídeos) en tiempo real entre dos o más usuarios. Dentro de éstos se incluyen servicios de correo electrónico como Gmail⁶² y Outlook⁶³

⁶² Gmail, <http://gmail.com/>

⁶³ Outlook, <http://outlook.com/>

(antiguamente conocido como Hotmail), y aplicaciones móviles como por ejemplo, WhatsApp⁶⁴ y Telegram⁶⁵.

Habitualmente, los servicios de mensajería instantánea contienen datos de contacto de los usuarios y sus conocidos y conversaciones, así como los contenidos intercambiados entre éstos. Dentro de las categorías establecidas, estos servicios pueden contener cualquiera de las establecidas, puesto que sus conversaciones pueden reflejar, por ejemplo, datos relacionados con sus activos y transacciones, competencias, información sobre ciudadanía y justicia, etc. No obstante, mayoritariamente estos servicios custodian características sociales de los usuarios (contactos y conversaciones).

Dado la sensibilidad de los datos que manejan estos servicios, puesto que son conversaciones privadas e íntimas entre dos o más personas, el acceso a dichos recursos es privado y en general, no existen datos públicos asociados a estos proveedores.

Algunos proveedores de servicios de mensajería instantánea permiten el acceso con el consentimiento del usuario a los recursos custodiados basándose en la exportación de datos en ficheros estáticos (Google (2015)), lo que permite migrar a otros servicios o a través de APIs (por ejemplo, Telegram que incluye distintas APIs para la creación de nuevos clientes).

Al igual que en el caso de las redes sociales, estas APIs pueden implementar mecanismos de petición (por ejemplo, para recuperar el historial de conversaciones de un usuario o su lista de contactos) o suscripción (para poder recibir notificaciones, por ejemplo, de la recepción de un nuevo mensaje del usuario). En definitiva, los mecanismos de recuperación asociados con los servicios de mensajería son ficheros estáticos de datos personales, peticiones y suscripciones.

6.2.2.3 Proveedores asociados a los dispositivos móviles

Los dispositivos móviles, como teléfonos inteligentes o tabletas, se han constituido como el primer medio de acceso a Internet y son ampliamente utilizados dentro de la sociedad como, por ejemplo, en el caso de España que están presentes en el 96.4% de los hogares (Instituto Nacional de Estadística – INE, 2014). Éstos contienen datos personales de los usuarios muy variados y realmente descriptivos: información de contactos, fotografías, localización, registro de llamadas y en general, todos aquellos datos generados por las aplicaciones móviles. Además, tal y como se detallará en 6.2.2.5 Proveedores de servicios y dispositivos de Internet de las cosas, los dispositivos móviles pueden ser el medio para obtener datos relacionados con el Internet de las cosas.

Dadas las categorías de recursos personales establecidos, principalmente los dispositivos móviles se caracterizan por custodiar datos de localización de los usuarios y características sociales (por ejemplo, datos de contacto), pero al igual que las redes sociales o servicios de mensajería, los contenidos almacenados en estos dispositivos pueden revelar otras características de los usuarios.

⁶⁴ WhatsApp, <https://web.whatsapp.com/>

⁶⁵ Telegram, <https://telegram.org/>

El acceso a los recursos custodiados por dispositivos es siempre privado y requiere que el usuario lleve a cabo determinadas tareas o conceda los permisos necesarios para su recuperación.

La mayoría de los dispositivos móviles permite a sus usuarios la descarga de datos personales almacenados a través de ficheros estáticos que pueden exportarse a un ordenador o enviarse a otros dispositivos o proveedores de servicios. Además, la recuperación de datos personales puede llevarse a cabo a través de aplicaciones móviles o programas que se ejecutan en dispositivos móviles inteligentes y que se encuentran disponibles a través de plataformas de distribución específicas de los sistemas operativos móviles (como Android⁶⁶, Apple iOS⁶⁷, BlackBerry OS⁶⁸ o Windows Phone⁶⁹). Cuando un usuario decide instalar una aplicación móvil, generalmente, tienen que dar acceso a los datos personales requeridos por ésta y dicha aplicación accede a los recursos solicitados a través de distintas APIs provistas por los proveedores de plataformas de distribución.

Dado que los dispositivos móviles gestionan características de localización de los usuarios, fuertemente asociadas a aspectos dinámicos y temporales de los mismos, éstos permiten la recuperación de su información a través de mecanismos de suscripción en la que se comunica el cambio de dicha información a entidades externas autorizadas.

En conclusión, los proveedores de dispositivos móviles permiten a los usuarios la recuperación o exportación de los recursos almacenados a través de ficheros estáticos. Además, en los dispositivos inteligentes, los proveedores de su sistema operativo incluyen distintos mecanismos, como, por ejemplo, los basados en APIs, que permiten llevar a cabo la recuperación por petición o suscripción de datos personales de los usuarios.

6.2.2.4 Brókeres de datos

Como se describió en 3.6.3.1 Brókeres de datos personales, los brókeres de datos son proveedores de servicios que recolectan datos personales de los usuarios de una gran variedad de fuentes de datos, generalmente sin el conocimiento y consentimiento del usuario (FTC, 2014).

La misión de estos proveedores es ofrecer los datos recolectados a terceras entidades a cambio de beneficios económicos u otros acuerdos comerciales. Los brókeres de datos acopian billones de elementos de datos personales de distinta naturaleza como datos demográficos, identificativos, sensibles, datos gubernamentales, de redes sociales, de propiedades (viviendas, automóviles, etc.), intereses y gustos generales, datos financieros, de viajes, compras y salud, entre otros. Es decir, cualquiera de las categorías de recursos personales establecidas.

La opacidad sobre los modelos de negocio de los brókeres de datos provoca una escasez de conocimiento en los mecanismos de recuperación de sus datos personales, puesto que como

⁶⁶ Android, <https://www.android.com/>

⁶⁷ Apple iOS, <https://www.apple.com/es/ios/>

⁶⁸ BlackBerry OS, <http://www.blackberryos.com/>

⁶⁹ Windows Phone, <https://www.windowsphone.com>

se ha señalado, los recursos que éstos poseen se intercambian mediante acuerdos comerciales.

6.2.2.5 Proveedores de servicios y dispositivos de Internet de las cosas

Cada vez más, los usuarios vinculan sus objetos personales a Internet para disfrutar de servicios que se apoyan en distintos tipos de sensores o los llamados *wearables* o tecnología que se lleva puesta (como relojes, gafas o pulseras inteligentes, entre otros) (IHS Electronics & Media, 2013). En este sentido, existen soluciones aplicadas en el área de la domótica y la interconexión de los elementos de la vivienda, sistemas de control biométrico que monitorizan las constantes vitales, alimentación o actividad física, servicios relacionados con la interconexión con el propio coche o muchos otros objetos de nuestra vida cotidiana.

La popularización y el uso extendido de los teléfonos móviles (6.2.2.3 Proveedores asociados a los dispositivos móviles) han convertido a éstos en una pieza clave para la prestación de servicios relacionados con el Internet de las cosas. Por una parte, los teléfonos móviles permiten establecer un medio adecuado y familiar para que los usuarios interactúen con otros dispositivos electrónicos cercanos o lejanos (por ejemplo, la televisión, la calefacción o distintos sensores del hogar de un usuario). Además, por otra parte, los teléfonos inteligentes permiten obtener información relevante sobre los usuarios como, por ejemplo, aquella relacionada con su salud, actividad física o localización. En relación a esto último, existen aplicaciones móviles (como, por ejemplo, Runtastic⁷⁰) que permiten monitorizar actividades deportivas de los usuarios o servicios específicos de las plataformas de dispositivos móviles (como Google Fit⁷¹ o Apple Salud⁷²) que además de monitorizar la actividad física de los usuarios en cualquier momento, permiten y habilitan la compartición de datos de salud entre distintas aplicaciones. Los ejemplos anteriores se enmarcan dentro de la denominada salud móvil (*mHealth*) que en los últimos años ha experimentado un gran crecimiento, existiendo cerca de cien mil aplicaciones relacionadas en distintas plataformas (Comisión Europea, 2014).

El aumento de estas interconexiones con el mundo real lleva asociado un aumento en el número de datos personales que se generan y por lo tanto, de características particulares de los usuarios. Dentro de los servicios relacionados con el Internet de las cosas, los recursos personales custodiados reflejan principalmente características biológicas e información de localización, aunque éstos pueden manifestar otras características como los activos que posee un usuario, sus características psicológicas, competencias o preferencias.

En general, los datos personales generados en el contexto de Internet de las cosas son recolectados y explotados por los propios fabricantes de los dispositivos y existe poca compartición de información entre las distintas líneas de productos o con entidades externas. Además, dada la sensibilidad que tienen asociadas las características biológicas que gestionan este tipo de servicios, normalmente, su acceso es privado y requiere el consentimiento expreso del usuario.

⁷⁰ Runtastic, <https://www.runtastic.com/es>

⁷¹ Google Fit, <https://fit.google.com/>

⁷² Apple Salud, <http://www.apple.com/es/ios/health/>

Cada vez más existen soluciones de exportación basadas en ficheros CSV (del inglés, *comma-separated values*) con distintos formatos establecidos por los fabricantes o recientemente, van apareciendo APIs propietarias que permiten la creación de nuevas aplicaciones basadas en dispositivos wearables (por ejemplo, la Google Glass Mirror API⁷³ para las gafas inteligentes de Google).

Los servicios relacionados con el Internet de las cosas incluyen datos temporales y dinámicos de los usuarios, como la localización, y por ello, la aplicación de mecanismos de suscripción son idóneos para el aprovechamiento de dicha información.

6.2.2.6 Proveedores de servicios específicos con acceso privado

Además de los ejemplos señalados anteriormente, existen multitud de proveedores de servicios especializados en un segmento concreto del usuario que pueden almacenar datos personales muy específicos y detallados de los usuarios como, por ejemplo, bancos, tiendas online, universidades, etc. De estos datos pueden inferirse características relevantes de los usuarios relacionadas con sus activos y transacciones, preferencias, competencias y en general, cualquiera de las categorías de recursos personales establecidas.

Los mecanismos de recuperación asociados con estos proveedores son muy variados, puesto que dependen del proveedor de servicios considerado. En términos generales, al tratarse de datos privados, los mecanismos de recuperación asociados serán ficheros estáticos, peticiones o suscripciones, ya que los programas araña están enfocados a la recuperación de datos públicos. Más adelante, en 6.4.2 Producción y recuperación de eventos personales de negocio en una arquitectura empresarial del ámbito bancario, se presenta un ejemplo de recuperación de datos personales en un proveedor de servicios telemáticos de banca. Dicha recuperación se basa en la producción de un nuevo tipo de datos personales que se generan dentro su arquitectura empresarial y que representan valor de negocio para la propia entidad u otras entidades externas (consumidores de modelos de usuario).

6.2.2.7 Proveedores de servicio específicos con acceso público

Al igual que en el punto anterior, existen multitud de servicios de acceso público en Internet que pueden incluir datos personales de los usuarios como por ejemplo blogs, páginas de opiniones, periódicos, boletines oficiales, etc. Esta información se puede asociar a un usuario específico fácilmente porque éste o sus objetos personales (por ejemplo, viviendas o coches) se identifican con pseudónimos o un identificador concreto asociado al usuario. Dicha información puede ser creada o generada directamente por el usuario al que refieren los datos o por terceras entidades.

Alguna de la información que puede extraerse de estas fuentes son las preferencias, opiniones, sentimientos de los usuarios, reputación online, etc. pero también información específica de registros públicos sobre la vivienda, juicios, multas, deudas, etc. que puede considerarse extremadamente sensible.

⁷³ Google Glass Mirror API, <https://developers.google.com/glass/develop/mirror/index>

Generalmente, los mecanismos para la recuperación de este tipo de información deben llevarlo a cabo la entidad solicitante y el mecanismo más utilizado son los programas araña, aunque también pueden recuperarse datos de estos proveedores a través de ficheros estáticos, mecanismos de petición o suscripción dependiendo de los servicios incorporados por el proveedor de servicios considerado.

6.2.3 Otras fuentes de datos y recursos disponibles

Además de las fuentes de datos citadas anteriormente, el proceso de modelado de usuario distribuido que lleva a cabo el marco de datos personales puede apoyarse en los recursos de distintos proveedores de servicios que no tienen por qué estar vinculados directamente con un usuario. La utilización de dichos recursos puede facilitar y mejorar la producción de nuevos modelos, por ejemplo, para inicializar o descubrir nuevas características de un usuario específico conforme a los datos generados por otros usuarios.

A diferencia del apartado anterior, los criterios de caracterización establecidos para el usuario y los proveedores de servicios con potencial de actuar como fuentes de datos deben actualizarse y ampliarse conforme a las siguientes consideraciones:

- Los **recursos custodiados** por las fuentes identificadas no tienen por qué incluir datos personales directamente relacionados con un usuario. En sentido, deberá indicarse qué recursos se han identificado.
- En general, se trata de identificar **recursos de acceso público** que están a disposición de terceras entidades para su reutilización y que pueden ser de utilidad en el marco de datos personales. También, se pueden considerar otros recursos de **acceso privado** que pueden estar custodiados por proveedores de servicios como los mencionados en el apartado anterior (redes sociales, servicios de mensajería, etc.).
- Deberán señalarse los **mecanismos para recuperación** de los recursos identificados, aunque principalmente éstos estarán basados en peticiones que realiza el sistema a entidades externas que permiten obtener los recursos en el momento preciso de su utilización o procesamiento.

A continuación, se describen dos tipos de proveedores (plataformas de datos abiertos y datos enlazados) y dos categorías de recursos (información de comunidad e información de contexto) que se han identificado de utilidad para el proceso de modelado de usuario que realiza el marco de datos personales.

6.2.3.1 Plataformas de datos abiertos

Las plataformas de datos abiertos (ver 3.6.3.3 Plataformas de datos abiertos) engloban un conjunto de datos públicamente accesibles que puede utilizarse y redistribuirse sin permisos especiales para su uso y reutilización, al igual que el código libre para los programas software. Estas plataformas están focalizadas en datos gubernamentales o científicos y en general, ofrecen datos globales sobre conjuntos de usuarios o zonas geográficas, sin datos que identifiquen directamente a un usuario concreto (Álvarez, 2014).

En general, los proveedores de estos datos son los que deciden el formato de representación e intercambio de los recursos custodiados. Así, típicamente se utilizan ficheros CSV, XML (del inglés, *eXtensible Markup Language*), RDF (del inglés, *Resource Description Framework*) e incluso, archivos Excel o PDF, que se recuperan a través de las plataformas establecidas normalmente a través de mecanismos de petición por parte de un solicitante.

Las plataformas de datos abiertos incluyen, o puede extraerse de éstas, información estadística que se puede utilizar para la definición de modelos de usuarios o para la actualización de éstos. Por ejemplo, gracias al estudio y análisis de datos abiertos se puede obtener información relevante sobre comportamientos estadísticos de los usuarios y estos datos se pueden aplicar en el proceso de inferencia de nuevos modelos.

En definitiva, estas plataformas incluyen datos públicos no relacionados directamente con el usuario que pueden utilizarse en el proceso de modelado de usuario y cuya recuperación debe llevarse a cabo a través de la descarga de ficheros estáticos o mecanismos de petición provistos por la plataforma considerada.

6.2.3.2 Plataformas de datos enlazados

Los datos enlazados (conocidos generalmente por su término en inglés, *linked data*) (Bizer et al., 2009) refieren al conjunto de mejores prácticas para publicar y conectar datos estructurados en la web, utilizando para ello estándares internacionales y semánticos del consorcio W3C. Éstos crean vínculos entre datos o recursos de distintas fuentes que permite relacionar distintos conceptos para realizar tareas de petición y razonamiento y que pueden enriquecer los contenidos de los modelos de usuario.

Por ejemplo, DBpedia⁷⁴ es una base de conocimiento que incluye un conjunto de datos estructurados extraídos de Wikipedia que interconecta conceptos identificados (como organizaciones, personajes, ciudades u otras localizaciones, películas, etc.) y sobre la cual se puede realizar distintas peticiones que pueden ser la fuente para la generación de nuevos modelos de usuario. Para ilustrar esto, si se persigue la definición de un modelo de usuario que incluya las posibles películas favoritas de un usuario, se pueden realizar peticiones a DBpedia para extraer las películas del actor o director favorito del usuario considerado.

Al igual que en el caso anterior, los recursos considerados son habitualmente de acceso público y su recuperación se lleva a cabo a través de mecanismos de petición y en concreto, mediante protocolos de tecnologías semánticas.

6.2.3.3 Información de la comunidad

Dentro del área de modelado de usuario, la información de comunidad refiere al conjunto global de los datos personales que un proveedor de servicios posee de todos sus usuarios. La utilización de dicha información en el proceso de modelado de usuario es una técnica denominada como **filtrado colaborativo** (Su y Khoshgoftaar, 2009) y es ampliamente utilizada por los sistemas recomendadores que se basan en la definición de modelos de usuarios específicos en función de los datos personales del resto de usuarios del sistema.

⁷⁴ DBpedia, <http://wiki.dbpedia.org/>

El filtrado colaborativo abarca distintas técnicas y algoritmos que permiten utilizar la información de comunidad, simplificando su complejidad, se pueden establecer tres fases principales para llevarlo a cabo:

1. Establecimiento de características comunes: se definen e identifican características básicas del usuario X sobre el que se quiere obtener un modelo de usuario. Por ejemplo, perfil demográfico (como edad y profesión) o preferencias musicales de un usuario concreto (cantantes o canciones favoritas).
2. Comparación: se comparan las características anteriores del usuario X con el resto de usuarios sobre los que se tiene datos (información de comunidad) para identificar usuarios similares a X.
3. Generación: se crea o actualiza el modelo de usuario para X a partir de los datos de los usuarios similares identificados en el punto dos. Por ejemplo, en un servicio de reproducción de música personalizada, si se identifica que dos usuarios X e Y tienen las mismas (o comparten bastantes similitudes en) preferencias musicales, es probable que a X le interesen otras canciones que Y escucha con asiduidad pero que X no ha escuchado nunca. De esta forma, las preferencias indicadas por el usuario Y se pueden asociar al usuario X dentro de su modelo de usuario.

En definitiva, la información de comunidad permite deducir o inferir características y compartimientos de los usuarios basándose en los datos de usuarios similares de un proveedor de servicios o una comunidad concreta. Conocer en aislado información sobre un usuario es muy beneficioso para llevar a cabo un objetivo específico pero la comparación o análisis de datos de otros usuarios puede enriquecer esta información e incluso, inicializar los modelos de los usuarios sobre los que no se tienen suficientes datos.

Los datos sobre la comunidad se pueden obtener de distintos proveedores de servicios, como, por ejemplo, redes sociales o cualquier proveedor de servicios especializado que contenga datos sobre una masa de usuarios. La caracterización de estas fuentes de datos ya se ha realizado dentro de 6.2.2 Proveedores de servicios como fuentes de datos personales, así, los datos podrán ser públicos o privados dependiendo de la información considerada y tendrán distintos mecanismos de recuperación asociados. A modo de ejemplo, se puede obtener información de comunidad de los perfiles públicos de redes sociales a través de los mecanismos de petición o suscripción incluidos por éstas. También, se pueden llegar a acuerdos para recuperar los datos anonimizados generados por distintos usuarios en un proveedor de servicios específico de acceso privado.

6.2.3.4 Información de contexto

El modelado de contexto (Brusilovsky y Millán, 2007) puede considerarse como una disciplina independiente al modelado de usuario, aunque ambos están tan interconectados que podemos encontrar modelos de usuario que incluyen información de contexto. En ocasiones, se consideran modelos de usuario de corta duración a aquellos que contienen mayoritariamente características de contexto del usuario. Algunos ejemplos de información de contexto son: la plataforma utilizada por el usuario (navegador, sistema operativo, dispositivo,

etc.), la localización, las condiciones ambientales (temperatura, luz o presión), condiciones personales (pulso, humor, actividad física, etc.), contexto social (por ejemplo, con quién se encuentra el usuario), las tareas que realiza (si está en el trabajo o en tiempo de ocio), etc.

Generalmente, estos datos se obtienen a través de los dispositivos utilizados por los usuarios (por ejemplo, dispositivos móviles y wearables), a través de proveedores de servicios donde se planifican eventos como redes sociales o calendarios, mediante la revelación de información de las comunidades de un usuario, etc. La caracterización de las fuentes de datos mencionadas conforme a la recuperación de dichos datos se ha realizado en 6.2.2 Proveedores de servicios como fuentes de datos personales. En términos generales, la información de contexto podrá ser pública y/o privada, dependiendo de la fuente de datos considerada y así, los mecanismos de recuperación asociados podrán ser cualquiera de los establecidos, aunque dada la temporalidad de la información de contexto los basados en suscripción son los más adecuados.

La información de contexto de los usuarios puede resultar de utilidad para múltiples consumidores de modelos de usuario del marco de datos personales, ya que permite conocer la situación actual de los usuarios o inferir información a corto plazo sobre éstos. Por ejemplo, los consumidores de modelos de usuario únicamente conociendo la locación de un usuario pueden ofrecer múltiples servicios, como la oferta de nuevos locales o actividades.

6.2.4 Resumen de las características de las fuentes de datos externas

En los apartados anteriores se han caracterizado tres tipos principales de fuentes de datos externas con potencial de abastecer al marco de datos personales de los recursos necesarios para llevar a cabo el modelado de usuario distribuido: el usuario del sistema, proveedores de servicios y otras fuentes y recursos disponibles. Las características halladas para cada una de ellas son:

- Los **usuarios** pueden proveer al sistema datos personales que crean o generan dentro del mismo o través de la importación de recursos que pueden obtener en otros proveedores de servicios. Para ello, el marco de datos personales debe ofrecer a los usuarios las herramientas necesarias que permitan llevarlo a cabo (por ejemplo, formularios) y que incentiven su participación.
- Los usuarios pueden delegar en el sistema la recuperación de datos personales que les refieren directamente y que están custodiados por distintos **proveedores de servicios**. Todos estos proveedores incluyen información de identificación sobre el usuario y cada uno de ellos se ha caracterizado conforme a: los recursos custodiados (información de identificación, características biológicas, localización, activos y transacciones, ciudadanía y justicia, preferencias, competencias, funciones, posiciones y roles, características sociales y características psicológicas), el tipo de acceso a dichos recursos (público y/o privado) y los mecanismos asociados para la recuperación de éstos (programas araña, ficheros estáticos, peticiones y/o suscripciones).

La siguiente tabla incluye un resumen con las características de cada proveedor de servicios analizado.

Tabla 6-2. Resumen de las características de los proveedores de servicios como fuente de datos.

Proveedores de servicios	Recursos custodiados	Tipo de acceso	Mecanismos de recuperación asociados
Redes sociales	Características sociales, psicológicas, preferencias y roles.	Público y privado, depende la configuración del usuario	Programas araña Ficheros estáticos Peticiones Suscripciones
Servicios de mensajería	Características sociales	Privado	Ficheros estáticos Peticiones Suscripciones
Dispositivos móviles	Localización y características sociales	Privado	Ficheros estáticos Peticiones Suscripciones
Brókeres de datos	No se puede determinar, opacidad de las actividades	Privado	No se puede determinar, habilitación por acuerdos comerciales
Internet de las cosas	Características biológicas y localización	Privado	Ficheros estáticos Peticiones Suscripciones
Proveedores específicos privados	Característicos del proveedor considerado	Privado	Variados, dependen del proveedor considerado
Proveedores específicos públicos	Característicos del proveedor considerado	Público	Programas araña u otros que dependen del proveedor considerado

- Existen **otras fuentes de datos**, así como **otros recursos** de especial relevancia que pueden custodiar los proveedores de servicios anteriores, que el marco de datos personales puede aprovechar para llevar a cabo el modelado de usuario distribuido. En sentido, se han identificado los siguientes:
 - **Plataformas de datos abiertos:** contienen datos públicos de los que se pueden extraer información estadística de distintos dominios. Éstos no refieren a un usuario específico y se pueden obtener a través de ficheros estáticos y mecanismos de recuperación basados en peticiones a la plataforma.
 - **Plataformas de datos enlazados:** ofrecen relaciones semánticas entre conceptos de distintos dominios que pueden aprovecharse para la producción de modelos de usuario. El acceso a estos datos generalmente es público y se

lleva a cabo a través de mecanismos de petición. En general, la información no refiere a un usuario concreto.

- **Información de comunidad:** refiere al conjunto de datos personales de otros usuarios del sistema o comunidades externas que puede aprovecharse para modelar a un usuario específico. Estos datos están custodiados por proveedores de servicios como, por ejemplo, redes sociales, servicios de mensajería o cualquier servicio que cuente con una masa de usuarios, incluido el propio marco de datos personales. El tipo de acceso a dichos datos depende de la fuente de datos considerada, así como los mecanismos de recuperación asociados.
- **Información de contexto:** dado que ésta puede obtenerse a través de distintas fuentes externas como, por ejemplo, otros usuarios, dispositivos móviles o proveedores de servicios, conviene reflejar dicha información de forma independiente. Cabe señalar que esta información tiene asociada una gran temporalidad y con ello, que su validez en el tiempo es de corto plazo. Además, dependiendo de la fuente de datos considerada, esta información podrá ser pública o privada y se podrán utilizar distintos mecanismos para su recuperación pero los basados en suscripción son idóneos dada su validez temporal.

6.3 Recuperación de recursos en el marco de datos personales

Como se ha podido observar en la sección anterior, el universo de datos que rodea a los usuarios es inmenso y cada uno de estos datos engloba una complejidad intrínseca asociada a aspectos particulares como el lugar de creación del dato, el grado de identificabilidad, la sensibilidad, el tipo de acceso o la validez temporal. Del análisis realizado se establecen los siguientes principios de diseño para la etapa de recuperación de datos:

- Cualquier dato personal asociado a una persona que sea identificado para su recuperación, independientemente de que éste sea público o privado, deberá ser considerado como un **dato sensible**, ya que su utilización puede tener impacto directo sobre la persona a la que aplica.

Aunque en un principio, los datos personales sensibles (3.1.1 Datos personales) aplican a, por ejemplo, ideologías políticas, religiosos, origen racial o tendencial sexuales, entre otras, un cambio de contexto de cualquier dato personal y aún más, el procesamiento y análisis de los mismos para obtener nueva información, puede entenderse por los usuarios como una invasión a su intimidad y privacidad (tal y como sostienen los marcos de evaluación de integridad contextual (Nissenbaum, 2004)).

En este orden de ideas, se deberá indicar al usuario los motivos por los cuales se recupera dicho dato y las connotaciones que esto supone y obtener su consentimiento y/o autorización para llevar a cabo la recuperación de los datos identificados.

- La recuperación de datos personales con **validez a corto plazo** o temporales como, por ejemplo, la localización, deberá llevarse a cabo a través de mecanismos de suscripción que permitan la recuperación de dicha información en el momento oportuno y cuando realmente tiene vigencia.

Tal y como se señaló en la introducción del presente capítulo, y tomando como referencia la Figura 6-1 que pone en correspondencia las etapas del proceso de modelado de usuario distribuido y los módulos de la arquitectura del marco de datos personales, el **módulo Recolector** es el encargado de llevar a cabo la etapa de recuperación de los recursos necesarios que permitirán generar modelos de usuario adaptados a las necesidades específicas de distintos consumidores de modelos, todo ello con la participación y consentimiento del usuario.

La Figura 6-2 muestra una visión detallada del módulo Recolector del marco de datos personales, su relación con el resto de módulos de la arquitectura de referencia (Gestor de usuarios, Registro, Generador y Almacén de recursos) y sus interacciones principales con las fuentes de datos externas identificadas (usuario, redes sociales, servicios de mensajería, etc. 6.2.4 Resumen de las características de las fuentes de datos externas). Aquí, los distintos proveedores de servicios con potencial de actuar como fuentes de datos externas al sistema se han agrupado conforme a si el acceso a sus datos es público o privado y se han reflejado los servicios que pueden incorporar para permitir la recuperación de sus recursos (Generador de ficheros, Servicios de peticiones o Servicios de suscripciones).

Conjuntamente a lo anterior, se indican las relaciones externas al marco de datos personales que puede mantener el usuario y que repercuten en la generación, y por lo tanto, en la recuperación, de datos personales. En primer lugar, los usuarios establecen relaciones con distintos proveedores de servicios (de acceso público y/o privado), utilizando sus servicios y por lo tanto, generando datos personales dentro de éstos. Además, el usuario puede tener distintos dispositivos (como, por ejemplo, teléfono móvil, las gafas o el coche) conectados y por consiguiente, éstos estarán gestionados por distintos proveedores de servicios de los que se pueden recuperar los datos generados por dichos dispositivos. Por otra parte, el usuario establece distintas relaciones (por ejemplo, de amistad o laborales) y similitudes con el resto de usuarios de distintas comunidades y éstos, a su vez, pueden generar datos relacionados con el usuario o consigo mismo que pueden utilizarse para el proceso de modelado de usuario.

Figura 6-2. Visión global de la recuperación de recursos en el marco de datos personales.

Antes de comenzar con la descripción de la Figura 6-2, cabe señalar que para facilitar la legibilidad de la misma se han omitido las relaciones que se establecen entre los componentes del Recolector (Recuperador de ficheros, Solicitador, Gestor de suscripciones y Escuchador) y los servicios que pueden incorporar los proveedores de servicios que contienen datos públicos (Generador de ficheros, Servicios de peticiones y Servicios de suscripciones). Las relaciones entre dichos componentes son análogas a las indicadas para los proveedores de servicios que contienen datos privados.

El módulo Recolector, y conforme a lo expuesto en el capítulo anterior (ver 5.2.2), está formado por distintos subcomponentes específicos que permiten llevar a cabo la recuperación de los datos de cada una de las fuentes externas consideradas. En concreto, éstos pueden agruparse como:

- **Recuperador de usuario.** En comunicación con el módulo Gestor de usuarios, este componente es el encargado de obtener los datos personales que aporta al sistema el usuario cuando actúa como fuente de datos. Así, este componente recupera los datos personales que un usuario puede aportar a través de distintos formularios provistos

por el sistema y recibe los ficheros estáticos de datos que el usuario puede obtener de distintos proveedores de servicios para incorporarlos en el marco de datos personales.

- **Recuperador de ficheros.** Este componente es el encargado de obtener datos para el proceso de modelado de usuario distribuido a través del mecanismo de recuperación basado en ficheros estáticos. Estos ficheros se pueden obtener a través de dos procedimientos principales:
 - El usuario puede aportar ficheros que previamente ha recuperado de un proveedor de servicios a través del módulo Gestor de Usuarios y éstos se redirigen al Recuperador de usuario que los envía al Recuperador de ficheros.
 - El marco de datos personales puede llegar a distintos acuerdos o implementar los mecanismos necesarios para obtener ficheros de datos de un proveedor de servicios (con acceso público o privado). Por ejemplo, este componente podría recuperar ficheros estáticos de distintas plataformas de datos abiertos.
- **Solicitador.** Éste permite llevar a cabo la recuperación de datos custodiados por distintos proveedores de servicios (públicos o privados) a través del mecanismo de recuperación basado en peticiones. Para ello se comunica con los servicios de peticiones que pueden incorporar los distintos proveedores de servicios. El Solicitador estará formado a su vez por distintos subcomponentes específicos, uno para cada proveedor considerado. A modo de ejemplo, un Solicitador puede incluir las llamadas a APIs provistas por las redes sociales para la obtención de la lista de amigos de un usuario o solicitudes a plataformas de datos enlazados para recuperar información sobre un determinado concepto.
- **Gestor de suscripciones.** Este componente activa y configura los mecanismos de suscripción que distintos proveedores de servicios con datos públicos o privados incorporan para poder recuperar los datos a medida que éstos se producen o actualizan en el proveedor origen. Como resultado del proceso que lleva a cabo este componente, se habilita un subcomponente específico del Escuchador.
- **Escuchador.** Este componente recibe los datos, o las notificaciones relacionadas con éstos, que los proveedores de servicios públicos o privados comunican al sistema en tiempo real o a medida que dichos datos se producen o actualizan dadas unas condiciones establecidas por el Gestor de suscripciones. Por ejemplo, retomando el caso de redes sociales que permiten la recuperación de sus recursos a través de suscripciones, se puede configurar a través del Gestor de suscripciones que se notifique al sistema cuando un usuario cambie su dirección de contacto y en este caso, el Escuchador es el responsable de recibir la notificación de que dicho cambio se ha producido.
- **Programas araña.** El módulo Recolector podrá incluir distintos programas araña para recuperar datos públicos de proveedores de servicios. Estos programas no requieren servicios específicos en los proveedores de servicios, únicamente que permitan el acceso a entidades externas, generalmente, a través de mecanismos de petición.

La activación y con ello, la ejecución, de cada uno de los componentes de recuperación descritos se realiza conforme a las preferencias e intereses de privacidad del usuario y por lo tanto, únicamente se lleva a cabo si éste da su consentimiento expreso o, en el caso de que aplique, conceda los permisos necesarios, tal y como se detallará en el siguiente capítulo.

Dadas las consideraciones anteriores, el módulo Recolector incluye un componente denominado como **Gestor de componentes** que coordina y orquesta el resto de elementos del módulo conforme a las decisiones del usuario. Para ello se comunica con el módulo Registro del marco de datos personales que almacena la información relacionada con los consentimientos y las autorizaciones concedidos por el usuario a través del módulo Gestor de Usuarios.

El Gestor de componentes también es el encargado de ofrecer una interfaz común a los módulos del marco de datos personales que requieran los recursos recuperados. En este sentido, los datos podrán ser demandados por el módulo Generador para producir modelos de usuario o dirigidos al Almacén de recursos para su almacenamiento. En el Capítulo 7 se entrará en detalle sobre estas relaciones.

En la siguiente sección se presentan las validaciones de distintos componentes del módulo Recolector que permiten llevar a cabo la recuperación de recursos privados en redes sociales, servicios de mensajería y un proveedor de servicios específicos del ámbito bancario a través de mecanismos basados en peticiones y suscripciones. En concreto, conforme al tipo de recursos establecidos en 6.2 Caracterización de las fuentes de datos externas, se ha recuperado información de identificación, datos sobre las transacciones o actividades realizadas por los usuarios y características sociales.

6.4 Validación de componentes específicos para la recuperación de datos

Tal y como se señaló en 5.3.1 Enfoque de validación global, en esta tesis doctoral se han considerado distintos escenarios que permiten validar los módulos principales, así como componentes específicos, del marco de datos personales. Para la etapa de recuperación de datos, se han establecido dos escenarios principales que se describen en esta sección y que permiten validar distintos mecanismos de recuperación basados en peticiones y suscripciones de datos privados.

A continuación, se describen los escenarios señalados, detallando los elementos y procedimientos que permiten la recuperación de datos personales para llevar a cabo el proceso de modelado de usuario distribuido.

6.4.1 Recuperación de datos personales en servicios de mensajería y redes sociales: el servicio de pago Posdata

En la presente investigación se ha desarrollado un servicio de pago directo entre personas, denominado como Posdata, que permite validar la recuperación de datos privados en

proveedores de servicios como redes sociales y servicios de mensajería a través de mecanismos de petición y suscripción. En concreto, los datos recuperados por este servicio son: información de identificación (nombre, apellidos y foto de perfil), características sociales de los usuarios (lista de amigos) y pagos realizados entre los usuarios del servicio.

Más allá de dicha validación, el servicio Posdata es una contribución en sí mismo, ya que supone un avance en el estado del arte de los servicios de pago personales. Por este motivo, en primer lugar, con el objeto de comprender el alcance del servicio Posdata, se incluye la descripción del contexto y motivación que han propiciado su desarrollo. Seguidamente, se presentan los trabajos relacionados y el posicionamiento del servicio respecto a éstos. Además, se indica el funcionamiento del servicio y por último, los detalles relacionados con la arquitectura del servicio y la validación de la etapa de recuperación de datos.

6.4.1.1 Contexto y motivación

El hecho de poder acceder a servicios financieros resulta habitual para gran parte de la sociedad de los países desarrollados, sin embargo esto no es así a nivel global. Más de 2.5 millardos de los adultos de la población mundial no utiliza servicios financieros (Chaia, et al., 2009) como, por ejemplo, tarjetas de crédito o débito, cheques, créditos, etc., bien porque no tienen acceso a servicios bancarios formales y por lo tanto, se encuentran financieramente excluidos, o bien porque aun teniendo alguna cuenta bancaria no aprovechan los servicios financieros electrónicos que tienen a su alcance (Ipsos MORI, 2013).

La exclusión financiera está profundamente vinculada a la situación de pobreza o exclusión social (Comisión Europea, 2008). En consecuencia, 2.2 millardos de personas, lo que supone un 62% de los adultos de la población mundial, no tienen acceso a servicios bancarios en África, Asia, América Latina y Oriente Medio. En los países de altos ingresos de la OECD la cifra no alcanza los 60 millones de personas (Chaia, et al., 2009).

Al mismo tiempo, incluso en los sectores con mayor índice de exclusión financiera, el uso de Internet está aumentando en los últimos años. Por ejemplo, en América Latina el número de usuarios de Internet sobrepasó los 300 millones en 2014 y la tendencia apunta a que va a seguir haciéndolo (The Statistics Portal, 2015). Además, dentro de estos usuarios de Internet, el 94% usa redes sociales, siendo la más popular Facebook con 180 millones de usuarios, seguida de Twitter, YouTube y LinkedIn (Simcott, 2014).

Existen causas sociales, como el envejecimiento de la población, que, unidas al desequilibrio tecnológico, aumentan la exclusión financiera y disminuyen la aceptación y utilización de servicios financieros electrónicos. Sin embargo, otros muchos factores, como por ejemplo, la desconfianza hacia los proveedores, el diseño inadecuado de los productos y servicios financieros, la preocupación por los costes y comisiones, etc. (Comisión Europea, 2013), que también contribuyen. Tal y como muestra el estudio (Ipsos MORI, 2013) desarrollado en el contexto europeo, el acceso a la tecnología y el uso de servicios telemáticos no está vinculado a una mayor utilización de servicios financieros. Así, en dicho estudio las personas no bancarizadas tienen una edad media de 40 años y la mayoría, un 96%, tienen acceso a algún dispositivo tecnológico como teléfono móvil u ordenador personal. Sin embargo, este sector de la población utiliza dichos dispositivos electrónicos para realizar tareas sencillas,

generalmente relacionadas con las redes sociales como por ejemplo, compartir mensajes y fotos con sus amigos de Facebook, y no acceden a servicios financieros electrónicos, ya que consideran estos últimos complicados.

La popularización de Internet y posteriormente, de los teléfonos inteligentes, han contribuido a la aparición de nuevos servicios financieros que, cuanto más sencillos e intuitivos resultan, más posibilidades tienen de ser adoptados por la población en general.

Consciente de lo anterior, el Center for Open Middleware (COM) lanzó a principios de 2014 un proyecto piloto, llamado POSDATA, cuyo objetivo era explorar las posibilidades que los servicios telemáticos y, en concreto, las redes sociales, ofrecen para la creación y desarrollo de nuevos servicios financieros dirigidos especialmente al sector de la población no bancarizada. Lo anterior ha permitido adquirir el conocimiento necesario para el desarrollo de servicios que se apoyan en los recursos de proveedores de servicios telemáticos y en concreto, en la recuperación de los datos personales que se generan en dichos proveedores.

6.4.1.2 Nuevos servicios de pago personales

Como ya se señaló en 4.3.2 Modelo de negocio para la promoción de nuevos servicios financieros, los servicios de pago telemáticos han avanzado en distintas direcciones y se han convertido en una de las áreas más competitivas del sector financiero. Así, podemos encontrar soluciones específicas para teléfonos móviles que permiten realizar compras en contextos físicos o distintos servicios que facilitan la realización de pagos en contextos telemáticos como Internet para, por ejemplo, la realización de compras o contrataciones de servicios.

En los últimos años, han ido apareciendo un conjunto de soluciones que permiten el intercambio de dinero directo entre personas sin la intervención aparente de una entidad bancaria y que ha sido denominada como pagos p2p.

Por una parte, distintos servicios de pago web, como PayPal o Google Wallet⁷⁵, han evolucionado para permitir la realización de pagos directos entre personas aunque originalmente fueron diseñados como mecanismo de pago entre personas y empresas minoristas. Estos proveedores emulan a la banca tradicional, actuando como intermediarios entre las personas, e imponen que los usuarios acudan a su sitio web cada vez que van a realizar un pago. Los usuarios tienen que crearse cuentas específicas en cada proveedor de servicios y vincular a dichas cuentas un método de pago bancario como una tarjeta o cuenta.

Por otra parte, han aparecido numerosas aplicaciones móviles para enviar dinero a otros usuarios, como, por ejemplo, Yaap Money⁷⁶, Venmo⁷⁷ o Cashually⁷⁸, que se basan en la comodidad de hacer pagos fáciles y rápidos a través de una aplicación instalada en el terminal móvil pero obligan a los usuarios a instalarse dicha aplicación. Al igual que en las soluciones web, estas soluciones no se integran en servicios ya existentes, los usuarios deben acudir a la

⁷⁵ Google Wallet, <https://wallet.google.com/>

⁷⁶ Yaap Money, <https://money.yaap.com/>

⁷⁷ Venmo, <https://venmo.com/>

⁷⁸ Cashually, <http://www.cashually.com/es/>

aplicación para poder acceder al servicio y asociar algún producto bancario para realizar los pagos.

Los dos enfoques anteriores se apoyan en productos bancarios, como tarjetas o cuentas bancarias, para hacer efectivos los pagos, lo cual obliga a sus usuarios a tener dichos productos bancarios contratados y además, requieren que los usuarios acudan a los servicios o aplicaciones específicas para realizar los pagos.

A diferencia de las soluciones anteriores, el servicio de pago Posdata se basa en integrar el pago entre personas en las propias funcionalidades que ofrecen los servicios de mensajería o redes sociales actuales, apoyándose en la utilización de la Tarjeta Universitaria Inteligente (TUI) emitida por el Banco Santander en más de 300 universidades de 11 países y con más de 7 millones de usuarios. Muchos de sus usuarios de la TUI proceden de regiones o segmentos de población no bancarizadas, lo cual permite aproximarse a dicho sector.

Cabe mencionar que en el momento de la realización del servicio de pago Posdata y posteriormente, comenzaron a aparecer algunas soluciones puntuales similares. Por ejemplo, BBVA Link⁷⁹ permite enviar dinero desde Facebook pero se basa en distintos formularios web integrados en una aplicación que se ejecuta en Facebook. También recientemente Facebook anunció para sus usuarios estadounidenses un servicio de pago integrado en su aplicación de chat⁸⁰.

6.4.1.3 Funcionamiento del servicio de pago Posdata

El servicio Posdata permite a sus usuarios realizar pagos a través de distintos canales de comunicación, como servicios de mensajería (por ejemplo, Telegram) o redes sociales (como Facebook o Twitter). Para ello se basa en la comunicación, mediante mensajes propios de cada canal, del envío de dinero entre los usuarios suscritos al servicio y en la realización final de los pagos mediante los saldos asociados a la tarjeta TUI.

Para darse de alta en el servicio, el usuario debe aportar la información relacionada con su TUI (identificador de la tarjeta). Seguidamente, éste tiene que activar al menos un canal de comunicación entre los disponibles (Facebook, Twitter o Telegram) para habilitar la realización de pagos a otros usuarios directamente a través de los canales deseados.

El servicio cuenta con una interfaz de usuario (Figura 6-3) que facilita la realización de dichas tareas y que, además, informa sobre sus términos de uso y políticas de privacidad, permite consultar los saldos asociados y últimas actividades realizadas.

⁷⁹ BBVA Link, https://www.bbva.cl/personas/bbva_link.jsp

⁸⁰ Facebook pagos en Messenger, <https://www.facebook.com/help/863171203733904/>

Figura 6-3. Interfaz de usuario de activación del servicio Posdata (San Miguel, et. al, 2015b)

Una vez que se ha activado el servicio Posdata y configurado los canales sociales deseados, los usuarios pueden comenzar a hacer pagos a sus amigos a través de la publicación de mensajes en los canales activos. Estos mensajes deben seguir un patrón predefinido que se le indica al usuario.

La siguiente figura muestra un ejemplo de un perfil de usuario de Facebook con los mensajes necesarios para realizar un pago dentro de la red social. La secuencia de dichos mensajes es la siguiente: el usuario Bob escribe una publicación para indicar que desea realizar un pago de 4.5 euros a su amiga Alice. Seguidamente, el servicio Posdata solicita un código de seguridad en un canal alternativo a Facebook que permite confirmar el pago y finalmente, informa al usuario sobre la ejecución de su orden. En el ejemplo de la figura, el pago se ha realizado con éxito.

Figura 6-4. Publicación de Facebook con la realización de un pago en el servicio Posdata (San Miguel, et. al, 2014c).

Además de lo anterior, el servicio Posdata está preparado para interactuar con los usuarios en otras situaciones. Por ejemplo, cuando un usuario escribe una publicación que no corresponde exactamente con el patrón establecido, Posdata responde con mensaje de la forma *“Lo siento, sólo entiendo mensajes de la forma: @Destino, te envío X € via @Posdata Service. Diríjase a [url] para más información”*. Aquí, [url] indica la dirección del servicio Posdata donde explica su funcionamiento a los usuarios.

6.4.1.4 Arquitectura del servicio y validación de la etapa de recuperación de datos

El servicio Posdata ha sido diseñado para permitir la realización de pagos p2p a través de distintos canales de comunicación existentes. La siguiente figura muestra la arquitectura del servicio: sus seis módulos (*User Manager*, *Registrar*, *Listeners*, *Dispatchers*, *User Data Store* y *Analyser*), las interacciones principales entre ellos y sus relaciones con los canales sociales de comunicación (Facebook, Twitter, Telegram u otros) y el Banco Santander (emisor de la tarjeta TUI).

Figura 6-5. Arquitectura del servicio Posdata.

Cada uno de los módulos del servicio Posdata mencionados valida parte de los módulos y componentes específicos que participan en la etapa de recuperación de datos descrita en 6.3 Recuperación de recursos en el marco de datos personales. En concreto:

- El módulo User Manager cubre distintas funcionalidades definidas para el Gestor de Usuarios del marco de datos personales. En concreto, en cuanto a la etapa de recuperación, incluye las interfaces necesarias que permiten a los usuarios dar su autorización para llevar a cabo la recuperación de datos custodiados en distintos proveedores de servicios de acceso privado (Facebook, Twitter o Telegram, entre otros).
- El modulo Registrar lleva a cabo funcionalidades del módulo Registro del marco de datos personales y del Gestor de componentes del módulo Recolector. Además, actúa como un Solicitador para recuperar recursos a través de mecanismos de recuperación basados en peticiones y como Gestor de suscripciones para activar la recuperación basada en suscripciones. Las funcionalidades concretas son:
 - Gestiona la información relacionada con las autorizaciones de acceso otorgadas por el usuario. Además, determina qué componentes de recuperación se activan dentro del servicio.
 - En función de los permisos concedidos por los usuarios, recupera datos personales de éstos a través de mecanismos de petición. En concreto, recupera información de identificación del usuario (nombre, apellidos y foto del perfil) y distintas características sociales como, por ejemplo, su lista de amigos.
 - Configura la suscripción a actualizaciones de datos personales de los usuarios y en particular, a los mensajes que éstos emiten para indicar que quieren realizar un pago.

- El módulo Listeners es equivalente al componente Escuchador definido para el módulo Recolector del marco de datos personales y por lo tanto, recibe avisos de los proveedores de servicios sobre la generación de nuevo contenido para llevar a cabo la recuperación basada en suscripciones. En particular, este contenido son los mensajes que el usuario escribe en el proveedor de servicios para indicar que quiere realizar un pago.

El resto de módulos incluidos en la arquitectura del servicio Posdata son propios de éste y cubren funcionalidades concretas del servicio. Simplificando la descripción de éstos, el módulo Dispatchers permite al servicio comunicarse con los usuarios a través de los distintos proveedores de servicios, el módulo Analyser es el encargado de procesar los datos personales generados por los usuarios (por ejemplo, los pagos) y el User Data Store almacena toda la información relacionada con el servicio.

La Figura 6-6 representa el diagrama de interacciones entre las entidades participantes (el usuario, el servicio de pago Posdata y un proveedor de mensajería o redes sociales externo, particularizado para Facebook) para llevar a cabo la recuperación de datos a través de mecanismos de petición y suscripción. Las relaciones 1 a 5 definen el procedimiento de autorización que permite llevar a cabo la recuperación de los datos dato1 y dato2. Seguidamente las interacciones 6 y 7 representan el mecanismo de recuperación basado en peticiones por el que se obtiene el dato1. Por último, los pasos 8 a 10 indican la recuperación del dato2 basada en el mecanismo de suscripción establecido por Facebook.

Figura 6-6. Proceso de recuperación de datos del servicio Posdata.

La implementación del servicio Posdata se ha realizado para la red social Facebook y para ello, se han utilizado las APIs que ésta ofrece. Dichas APIs son similares a las ofrecidas por otros proveedores de redes sociales, como Twitter, y por lo tanto, puede extrapolarse la implementación a otros servicios externos fácilmente. Las APIs utilizadas para la validación han sido:

- El producto de Inicio de Sesión de Facebook que permite obtener la autorización del usuario para llevar a cabo la recuperación de datos personales. Éste se basa en una variante del protocolo OAuth 2.0.
- La Graph API de Facebook para obtener datos personales de los usuarios a través de peticiones siguiendo un enfoque REST.
- La funcionalidad *Real Time Updates* del Graph API de Facebook para suscribirse a la notificación de nuevas publicaciones de los usuarios relacionadas con los pagos que éstos desean enviar a través de la red social. En el caso particular de Facebook, estas notificaciones indican que se han producido nuevas publicaciones pero deben realizarse peticiones REST a través del Graph API que permitan recuperar el contenido de la publicación del usuario.

Por último, cabe señalar en cuanto al desarrollo del servicio Posdata que éste se ha implementado como un servicio web basado en tecnologías de servidor Apache, utilizando JavaScript y PHP como lenguajes de desarrollo. Para el primer prototipo desarrollado, se han simulado las funcionalidades y comunicaciones con el Banco Santander. Además, el módulo Dispatcher se apoya también en el Graph API de Facebook para comunicarse con los usuarios a través de la red social. Por su parte, el módulo Analyser incluye distintos *scripts* que detectan el significado de las publicaciones de los usuarios, apoyándose en expresiones regulares que determinan si la sintaxis es adecuada en conjunto con los metadatos recibidos de la red social. Finalmente, el almacenamiento de los datos de usuario se basa en un sistema relacional administrado con MySQL.

6.4.2 Producción y recuperación de eventos personales de negocio en una arquitectura empresarial del ámbito bancario

El servicio de pago Posdata ha permitido validar la recuperación de datos personales custodiados por proveedores de servicios externos (servicios de mensajería y de redes sociales) a través de mecanismos de petición y suscripción que fundamentalmente se llevan a cabo a través de las APIs provistas por dichos servicios. Conjuntamente, en el contexto de esta investigación, se ha analizado la arquitectura empresarial de un proveedor de servicios bancarios que requiere evolucionar hacia soluciones dinámicas basadas en la producción y recuperación de un nuevo tipo de datos personales, que ha sido denominado como **eventos personales de negocio**. La recuperación de estos datos por parte del marco de datos personales se realiza a través de mecanismos de suscripción y requiere su identificación y producción dentro de la arquitectura empresarial del proveedor de servicios que actuará como fuente de datos externa al sistema.

Tal y como se señaló en 4.3.2 Modelo de negocio para la promoción de nuevos servicios financieros, los bancos, al igual que muchas otras organizaciones empresariales actuales, se encuentran inmersos en un entorno de gran competitividad global que precisa de soluciones específicas para adaptarse a los nuevos tiempos. El término arquitectura de empresa o arquitectura empresarial se acuñó a mediados de los años 90 para permitir a las

organizaciones una alineación de sus objetivos y procesos de negocio, sistemas de información, datos e infraestructuras tecnológicas y así, hacer frente a los desafíos del nuevo contexto (Arango Serna, 2010).

Dentro de una arquitectura empresarial, se pueden identificar multitud de tipos de eventos o cambios de estado dentro de ésta que merecen atención en otras entidades para su uso y reutilización. Por ejemplo, existen eventos relacionados con información tecnológica de bajo nivel (como datos generados por distintos sensores o la infraestructura de red), con los procesos software del proveedor (como las llamadas a distintos métodos o las ejecuciones de excepciones en el código), con las actividades que realizan los usuarios (incluyendo dentro de éstos las acciones realizadas por los clientes y los trabajadores), etc.

Los trabajos presentados en este escenario de validación se basan en eventos de negocio que repercuten directamente en el desarrollo económico de la empresa y pueden ser utilizados para la mejora de servicios existentes o para la creación de otros nuevos (tal y como se detallará en el siguiente capítulo). Estos eventos se basan y definen a partir de los datos personales que generan los clientes de un proveedor de servicios bancario y por ello, se han denominado como eventos personales de negocio.

Los eventos personales de negocio reflejan información de contexto de los clientes de la entidad, en particular, las acciones que realizan a través de los servicios y aplicaciones bancarios ofrecidos, y tienen impacto directo en el negocio. Por ejemplo, eventos personales de negocio son aquellos que indican si un usuario determinado ha contratado o cancelado un nuevo servicio financiero (como un depósito o un plan de pensiones), retirado efectivo, realizado o recibido una transferencia, consultado un anuncio publicado en un servicio, etc. en un determinado momento.

A continuación, en primer lugar, y con el objeto de poner en contexto al lector, se describe la arquitectura empresarial del ámbito bancario que se ha analizado. Seguidamente, se incluyen las soluciones realizadas que permiten llevar a cabo la producción de eventos personales de negocio y su recuperación en el marco de datos personales. En concreto, se presenta una propuesta para la definición de los eventos que pueden producirse dentro de la arquitectura empresarial y se establece una arquitectura de referencia para la producción y recuperación del nuevo tipo de datos personales identificado.

6.4.2.1 Descripción de la arquitectura empresarial BankSphere

En el contexto de esta investigación se ha analizado una arquitectura empresarial del Banco Santander creada por su división tecnológica ISBAN, denominada como BankSphere (BKS). Esta arquitectura requiere evolucionar hacia el paradigma de arquitecturas dirigidas a eventos (3.3.4.1 Arquitectura dirigida a eventos) para permitir la generación y recuperación de eventos personales de negocio.

BKS es un conjunto de herramientas de desarrollo que permite la creación de aplicaciones bancarias de forma sencilla, rápida y eficaz. Ésta ofrece un entorno de diseño integrado con la

plataforma de desarrollo Eclipse⁸¹ y el despliegue y entorno de ejecución se basa en soluciones Java y tecnologías web.

La arquitectura está diseñada bajo dos premisas fundamentales: permitir la reutilización de componentes software y facilitar una programación rápida y sencilla. La primera se logra gracias a la aplicación del paradigma de arquitectura orientada a servicios, donde distintos componentes software se desarrollan y exponen para ser reutilizadas por terceros. Por su parte, la segunda premisa ha propiciado el desarrollo de un entorno de programación visual que permite el diseño de aplicaciones a través de distintos símbolos gráficos que representan los componentes software de la aplicación y que abstraen de los detalles de implementación concretos a los programadores.

Simplificando el desarrollo de una aplicación BKS, ésta consiste en la definición de distintos diagramas de estados, formados por flujos de presentación y de negocio, que reutilizan componentes software previamente implementados. En general, una aplicación está formada por un flujo de presentación principal que realiza distintas llamadas a flujos de negocio que, a su vez, pueden disparar la ejecución de operaciones y servicios de sistema (conocidos por sus términos en inglés, *backend*). Los flujos de negocio se exponen a través de un elemento denominado como fachada o *facade* dentro de la arquitectura BKS.

Una aplicación BKS se implementa como una aplicación *Java Enterprise Edition* empaquetada (EAR), cuenta con un módulo de presentación web (WAR) y varios módulos de negocio (*Enterprise JavaBeans* – EJBs). Su ejecución se lleva a cabo en un entorno también provisto dentro de la arquitectura empresarial tal y como muestra la siguiente figura.

Figura 6-7. Entorno de ejecución de BankSphere (San Miguel, et al., 2014b).

La secuencia de ejecución de una aplicación BKS es la siguiente: en primer lugar, cuando se detecta la petición de una aplicación concreta, se redirige dicha petición a un contenedor de operaciones (*operation container*) que invoca una serie de operaciones iniciales y estados definidos por el flujo de presentación de la aplicación (en la figura, *presentation flow* – PF). El paso por los distintos estados definidos por el flujo de presentación puede desencadenar la llamada a flujos de negocio. Esta llamada se realiza a través de una fachada común (en la figura, *facade*) que implica la llamada a operaciones de sistema (backend) y otros

⁸¹ Eclipse, <https://eclipse.org/>

componentes externos definidos por la lógica de negocio. La llamada a la aplicación BKS finaliza cuando se ejecutan todos los estados de presentación y con ello, se devuelve el control al usuario (de forma general, éstos serán clientes y trabajadores).

En los siguientes apartados se describen las propuestas y validaciones que permiten evolucionar la arquitectura BKS presentada para producir y recuperar eventos personales de negocio. En primer lugar se presenta una definición genérica de los eventos que pueden producirse dentro de la arquitectura y seguidamente, los detalles relacionados con la etapa de recuperación por parte del marco de datos personales.

6.4.2.2 Definición genérica de eventos

Tal y como se indicó en el estado del arte (3.3.4.1 Arquitectura dirigida a eventos), la definición de eventos es uno de los retos más importantes de una arquitectura EDA, puesto que éstos son la pieza fundamental sobre la que se sustenta.

Actualmente, se pueden encontrar múltiples alternativas para la definición de eventos y no existe un consenso sobre los formatos o semánticas a utilizar. Por ejemplo, existen soluciones basadas en objetos Java, lenguajes XML u otros lenguajes de intercambio de datos como JSON (IETF, 2014) o el Google Protocol Buffers⁸². Ante este contexto, distintos estudios como, por ejemplo, los incluidos en (Michelson, 2006) y (Etzion y Niblett, 2010) coinciden en la necesidad de definir un formato único que permita la interoperabilidad de los distintos eventos y señalan un conjunto de directrices para la definición de eventos.

En el contexto de este trabajo, se han seguido las directrices extraídas de las publicaciones citadas anteriormente y se ha propuesto una definición para los distintos eventos que pueden producirse dentro de la arquitectura empresarial BKS. Dicha definición permite expresar cualquier tipo de evento, independientemente de su origen o naturaleza como, por ejemplo, aquellos relacionados con datos de sensores o infraestructuras tecnológicas, además de los de eventos personales de negocio objeto de recuperación en el escenario de validación presentado.

Dadas las consideraciones anteriores, se ha propuesto una definición genérica que permite representar cualquier tipo de evento y que contiene dos partes diferenciadas:

- **Cabecera:** incluye información y metadatos comunes a todos los eventos que pueden darse dentro de la arquitectura empresarial. En particular, se han identificado tres elementos dentro de ésta:
 - **Tipo de evento (*eventType*):** indica el tipo de evento de acuerdo con la jerarquía de eventos establecida por el Grupo Santander. De esta forma, el tipo de evento viene determinado por su categoría o dominio en el que aplica dentro de la arquitectura (por ejemplo, de negocio o hardware), el área de negocio donde se origina (servicios de transferencias online, hipotecas, etc.) y el tipo de evento específico dentro del área de negocio (por ejemplo, si se ha contratado o cancelado un determinado servicio).

⁸² Google Developers, Protocol Buffers, <https://developers.google.com/protocol-buffers/>

Centrándonos en el caso de eventos personales de negocio, como su propio nombre indica, todos éstos están dentro de la categoría “negocio”. Dentro de ésta existen múltiples áreas de negocio, tantos como servicios financieros ofrecidos a los usuarios, por ejemplo, servicios de transferencias, hipotecas, depósitos, planes de pensiones, publicidad, etc. Además, en cada una de estas áreas existen distintos eventos personales que reflejan las acciones que pueden realizar los usuarios (como contratación, baja, visualización, etc.).

- **Tiempo de creación** (*createdTime*): indica el momento preciso en la que se generó el evento.
- **Productor** (*createdBy*): contiene el identificador del servicio o aplicación concreta que generó el evento.
- **Cuerpo**: contiene datos específicos del tipo de evento generado. Por ejemplo, para un evento personal de negocio que indique que un usuario ha realizado una transferencia, el cuerpo incluye información relacionada con dicho evento como la identificación del usuario que realizó la transferencia, la sesión o dirección IP asociada con dicho usuario, la cuenta destino y origen de la transferencia, el importe enviado y el concepto o descripción asociada a la transferencia emitida.

En la propuesta realizada, se ha definido la estructura general de un evento (cabecera y cuerpo) a través de un *XML Schema Definition* (XSD) que hace referencia a otros XSD que contienen la definición asociada a cada tipo de evento de forma independiente. Esto permite que actores especializados dentro del Grupo Santander realicen la definición de los tipos de eventos específicos dentro de las distintas unidades y departamentos del grupo.

La Figura 6-8 representa gráficamente la definición de eventos propuesta para arquitectura empresarial BKS. Los tres primeros elementos (*eventType*, *createdTime* y *createdBy*) conforman la cabecera del evento. Por su parte, el elemento *logRecordAppData* delimita el cuerpo del evento y su contenido viene determinado por XSDs independientes que definen los tipos de eventos concretos.

Figura 6-8. Definición de los eventos de una arquitectura empresarial (San Miguel, et al., 2015d).

Una vez establecida la definición de eventos, en el siguiente apartado se describe una arquitectura de referencia propuesta para la producción de eventos personales dentro de la arquitectura BKS y su recuperación por parte del marco de datos personales.

6.4.2.3 Arquitectura de referencia para la recuperación de eventos personales de negocio

Conforme a lo expuesto en 6.4.2.1 Descripción de la arquitectura empresarial BankSphere, dicha arquitectura no permite la producción y por lo tanto, recuperación de eventos personales de negocio. Para tal efecto, en el contexto de este trabajo, se ha propuesto una arquitectura de referencia que identifica distintos métodos para llevar a cabo la producción de dichos eventos y los elementos necesarios para la recuperación de éstos por parte del marco de datos personales. Dicha arquitectura se sustenta en el establecimiento de una arquitectura EDA (ver 3.3.4.1), basada en la comunicación de eventos y tres niveles principales: productores, canal y consumidores.

La Figura 6-9 incluye la propuesta de arquitectura de referencia para la producción y recuperación de eventos personales de negocio. En la parte superior de la figura se observan los detalles relacionados con la producción conforme a la arquitectura BKS descrita en 6.4.2.1. En la parte inferior, se muestra la recuperación de dichos eventos en el marco de datos personales, identificando los módulos de la arquitectura que participan, así como los componentes de específicos del módulo Recolector, tal y como se estableció en la sección 6.3 Recuperación de recursos en el marco de datos personales. Asimismo, en la parte derecha de la figura se incluye la representación de los niveles de una arquitectura EDA genérica y su correspondencia con la arquitectura de referencia propuesta.

Figura 6-9. Arquitectura de referencia para la producción y recuperación de eventos personales de negocio en BankSphere.

En la figura se observa que las aplicaciones BKS actúan como fuentes de eventos personales de eventos y por lo tanto, toman el papel de productores dentro de la arquitectura EDA establecida. Los eventos producidos se transmiten a través de un canal de comunicación al marco de datos personales y en concreto, a su módulo Recolector, y así, el sistema actúa como un consumidor de los eventos producidos. A continuación, se describen los tres niveles principales de la arquitectura de referencia establecida (productores, canal y consumidores) y finalmente, se incluyen los detalles de implementación relacionados.

Producción de eventos personales de negocio

Tal y como se ha indicado, las aplicaciones BKS actúan como productores de eventos dentro de la arquitectura de referencia establecida y por consiguiente, deben llevar a cabo la identificación y producción de eventos personales de negocio.

En el contexto de este trabajo, se han propuesto dos métodos principales que permiten producir eventos personales de negocio dentro de las aplicaciones BKS:

- **Método explícito.** En este método, los programadores o desarrolladores de las aplicaciones BKS deben identificar e incorporar la producción de eventos personales de negocio dentro de la lógica de negocio de la aplicación, lo cual requiere obtener datos específicos relacionados con el contexto del usuario y las instancias de los eventos.

El método explícito tiene varios inconvenientes asociados. En primer lugar, los programadores tienen que adquirir nuevas responsabilidades y habilidades relacionadas con distintas áreas de negocio y esto, en términos generales, se aleja de las labores técnicas asociadas con su perfil profesional. Además, dado que la incorporación de eventos a través de este método requiere que se modifiquen aplicaciones BKS en producción, se pueden alterar aplicaciones estables, lo cual entraña riesgos en la provisión de servicios existentes.

- **Método implícito.** Aquí, la incorporación de la producción de eventos personales de negocio se lleva a cabo de forma casi transparente a los programadores de aplicaciones BKS. Para ello, la producción de eventos se asocia a las llamadas que una aplicación BKS realiza a la lógica de negocio, a través de la fachada común, tal y como se describió en 6.4.2.1.

El mayor inconveniente asociado con este método es que las llamadas que se realizan a la lógica de negocio a través de la fachada pueden no corresponder directamente con eventos personales de negocio. Para resolver esto, se ha propuesto incorporar un preprocesador de eventos que en base a las distintas llamadas realizadas a la lógica de negocio pueda generar eventos personales de negocio. La incorporación de este elemento requiere llevar a cabo un análisis sobre cada uno de los componentes funcionales provistos por la arquitectura BKS.

Los dos métodos propuestos pueden coexistir dentro de la arquitectura empresarial BKS, sin embargo, con el objeto de reducir al mínimo los riesgos asociados a la modificación de

aplicaciones BKS estables, se ha propuesto la utilización del método explícito para nuevas aplicaciones BKS y el uso del método implícito para aplicaciones existentes.

Transporte y comunicación de eventos personales de negocio

Una vez que se han identificado y producido los eventos personales de negocio dentro de la arquitectura empresarial, éstos se comunican y redirigen a distintos consumidores asociados a través de un canal de comunicación.

Aunque dicho canal no es objeto de estudio directo de la presente tesis doctoral, se ha propuesto la incorporación de un MOM (ver 3.3.4.1) que permite el envío y la recepción de mensajes a distintas entidades y en concreto, a través de un modelo de mensajería de publicación/suscripción que permitirá la recepción de eventos a más de una entidad interesada.

Consumo y recuperación de eventos personales de negocio

Dentro de la arquitectura de referencia propuesta (ver Figura 6-9), se ha considerado un único consumidor de eventos personales de negocio: el marco de datos personales, aunque pueden considerarse otras entidades o sistemas propios de la arquitectura BKS.

La recuperación de eventos personales de negocio por parte del marco de datos personales se lleva a cabo a través de mecanismos de suscripción y por lo tanto, tal y como se indicó en sección 6.3 Recuperación de recursos en el marco de datos personales, participan los siguientes módulos y componentes específicos de éste:

- **Gestor de usuarios:** permite a los usuarios del marco de datos personales, que a su vez utilizan los servicios y aplicaciones bancarias provistos por la arquitectura BKS, configurar y autorizar la recuperación de eventos personales de negocio que les refieren directamente.
- **Registro:** sobre la base de las autorizaciones realizadas por los usuarios, se comunica con el módulo Recolector del marco de datos personales, y en particular, con su componente Gestor de componentes, para activar o desactivar la recuperación de eventos personales de negocio.
- **Recolector:** tal y como se detalló en la sección 6.3, los distintos componentes que participan en el escenario de validación presentado son:
 - **Gestor de componentes:** es el encargado de activar la ejecución de los mecanismos de recuperación concretos que el marco de datos personales conforme a las preferencias del usuario.
 - **Gestor de suscripciones:** configura a través de los servicios de suscripción provistos por la arquitectura BKS la suscripción a distintos eventos personales de negocio. Como resultado de este proceso, se activan distintos subcomponentes del Escuchador que recibirán los eventos personales de negocio.

- **Escuchador:** estará formado por distintos escuchadores específicos que recibirán los distintos tipos de eventos personales de negocio a los que el marco de datos personales está suscrito. En la Figura 6-9 se han representado dos escuchadores (Escuchador 1 y Escuchador 2) que reciben los eventos producidos por las aplicaciones BKS a través de métodos implícitos y explícitos respectivamente.

Detalles de implementación

En el contexto de esta investigación se ha desarrollado un prototipo de experimentación que ha permitido validar la arquitectura de referencia propuesta.

En concreto, se han desarrollado distintos productores de eventos a través de tecnología Java que simulan aplicaciones BKS y que producen eventos personales de negocio conforme a la definición de eventos propuesta en 6.4.2.2. De esta forma, las instancias de los eventos producidos se comunican utilizando el lenguaje XML a través de un canal de comunicación.

Para el canal de comunicación se ha utilizado la herramienta de código libre RabbitMQ que cuenta con una amplia comunidad de desarrolladores y que además ofrece la posibilidad de adquirir servicios de licencias, consultoría o soporte, lo cual resultaba de utilidad dentro del Grupo Santander y era un requisito del proyecto en el que se desarrollaron los trabajos. RabbitMQ implementa el protocolo AMQP y permite la configuración de modelos de mensajería publicación/suscripción.

Los módulos de la arquitectura del marco de datos personales y los componentes específicos de éstos, se basan principalmente en Java como lenguaje de desarrollo, un servidor Apache, tecnologías HTML y JavaScript para la interfaces de usuario y librerías y funcionalidades ofrecidas por la herramienta RabbitMQ.

Conjuntamente a todo lo anterior, el prototipo desarrollado ha permitido validar requisitos no funcionales de los componentes seleccionados, relacionados éstos en su mayoría con el rendimiento de la herramienta RabbitMQ. En (San Miguel, et al., 2015d) y (Montero, 2014) se puede ampliar los trabajos mencionados.

6.5 Resumen y contribuciones del capítulo

En este capítulo se han presentado las contribuciones originales de la tesis doctoral, así como las validaciones realizadas, respecto a uno de los pilares fundamentales del marco de datos personales: la recuperación de recursos.

La recuperación de recursos supone la primera etapa del proceso de modelado de usuario distribuido que lleva a cabo el marco de datos personales y permite identificar y obtener los recursos necesarios para generar modelos de usuario, todo ello considerando la participación y control del usuario. Para lo anterior, se ha profundizado en el diseño del módulo Recolector y en las interacciones que éste establece con el resto de módulos del sistema para llevar a cabo la recuperación de recursos.

En concreto, las contribuciones incluidas en el capítulo son:

- La identificación y caracterización de las fuentes de datos externas con potencial de abastecer al marco de datos personales de los recursos necesarios. Las contribuciones originales a este respecto son: la consideración del usuario como fuente de datos y la identificación de un nuevo tipo de datos personales (eventos personales de negocio) que puede originarse dentro de una arquitectura empresarial.
- Propuesta de un modelo global para la recuperación de recursos heterogéneos distribuidos en distintas fuentes de datos.
- Diseño y desarrollo de un nuevo servicio de pago personal que a su vez ha permitido validar la recuperación de recursos personales en distintos proveedores de servicios de redes sociales y servicios de mensajería.
- Definición de un nuevo tipo de datos personales identificado en el contexto de esta investigación que ha sido denominado como eventos personales de negocio.

Cada una de las contribuciones señaladas ha sido fruto de la colaboración y participación en distintos proyectos de investigación en los que la autora de esta tesis ha participado. En particular, las contribuciones y las validaciones presentadas se han llevado a cabo en el marco de los proyectos MOET y POSDATA del Center for Open Middleware. Asimismo, cabe señalar, que dentro del proyecto Segur@ se identificaron y caracterizaron distintas fuentes de datos personales relacionadas con los dispositivos electrónicos de los usuarios. También, en el proyecto BUSCAMEDIA se experimentó y validaron dos componentes basados en programas araña que obtienen recursos de fuentes de datos externas. Estos últimos trabajos se incorporaron en distintos entregables de los proyectos señalados.

Todo lo anterior ha propiciado la publicación de los siguientes artículos:

- *Integrating Business Information Streams in a Core Banking Architecture: A Practical Experience* (San Miguel, et al., 2015d) incluido como capítulo del libro *Enterprise Information Systems*. Este artículo es fruto de la selección del artículo (San Miguel, et al., 2014b) como mejor del *16th International Conference on Enterprise Information Systems (ICEIS 2014)* en el área de arquitectura empresarial. Incluye los trabajos relacionados con la arquitectura empresarial presentada y la generación y recuperación del nuevo tipo de datos personales identificado (eventos personales de negocio), así como distintas validaciones relacionadas con el rendimiento de las herramientas seleccionadas.
- *Approaching Unbanked People: A Person-to-Person Payment Application on Facebook* (San Miguel, et al., 2015b) presentado en el VII Congreso Iberoamericano de Telemática (CITA2015). Éste incluye una descripción del servicio de pago personal desarrollado y los detalles para la recuperación de recursos que lleva a cabo y que ha permitido validar distintos componentes de la arquitectura.
- *Creating and Modelling Personal Socio-Economic Networks in On-Line Banking* (San Miguel, et al., 2014c) presentado en el *7th International Workshop on Context-*

Awareness and Personalization in Cloud and Service Computing (PCS 2014) donde fue seleccionado entre los mejores del congreso para la incorporación del artículo como capítulo de libro. Incluye aspectos relacionados con la recuperación de datos personales y su almacenamiento mediante tecnologías semánticas.

- *Evolving a Core Banking Enterprise Architecture: Leveraging Business Events Exploitation* (San Miguel, et al., 2014b) presentado en la *16th International Conference on Enterprise Information Systems (ICEIS 2014)* donde obtuvo el premio a mejor artículo del congreso en el área de arquitectura empresarial. Incluye los detalles técnicos de la arquitectura empresarial analiza en el contexto de investigación.

En el siguiente capítulo se presentarán las contribuciones originales de la tesis respecto a la generación de modelos de usuario a partir de recursos heterogéneos y distribuidos que se han identificado en este capítulo. Además, se describirán dos validaciones para la generación y aplicación de modelos de usuario que se basan en la reutilización de los recursos recuperados en las validaciones de este capítulo: aquellos obtenidos por el servicio de pago Posdata y los identificados en la arquitectura empresarial BKS.

Capítulo 7.

Generación y aplicación de modelos de usuario

En el capítulo anterior se presentaron las contribuciones y validaciones relacionadas con la primera etapa del proceso de modelado de usuario distribuido que permite obtener los recursos necesarios para producir modelos de usuario en el marco de datos personales. El presente capítulo aborda los retos relacionados con el resto de etapas de dicho proceso: generación y aplicación de modelos de usuario.

Por una parte, la etapa de **generación de modelos de usuario** engloba todas las acciones que deben realizarse para producir modelos a partir de los recursos identificados en la etapa de recuperación de datos con el objeto de cubrir las necesidades de negocio específicas de terceras entidades (consumidores de modelos de usuario). A diferencia del proceso de modelado de usuario tradicional (3.2 Proceso de modelado de usuario), donde típicamente un proveedor de servicios almacena los recursos recolectados de sus usuarios y procesa éstos para generar modelos de usuario, en el marco de datos personales deben considerarse las interrelaciones que se establecen con las distintas entidades externas (fuentes de datos y consumidores) y que repercuten en el establecimiento de distintos patrones para la generación de modelos de usuario que se describirán en la sección 7.2 Patrones para la generación de modelos de usuario.

Por otra parte, la aplicación de los modelos de usuario es la última etapa del proceso de modelado de usuario y consiste en el uso que se les da a los modelos generados. Aunque en un principio esta fase es ajena al marco de datos personales, puesto que cada consumidor de modelos de usuario lleva a cabo la aplicación de éstos, el sistema debe incorporar distintos mecanismos que permitan recuperar los modelos deseados y en el momento oportuno. Así, para el proceso de modelado de usuario distribuido, aparece una nueva etapa que ha sido denominada como **recuperación de modelos de usuario**, cuyos mecanismos serán muy similares a los utilizados en la etapa de recuperación de datos, y que repercutirá en la generación de modelos, ya que determina su inicialización y los recursos necesarios.

Posteriormente a la recuperación de modelos de usuario, le sigue la etapa de **aplicación de dichos modelos** que llevará a cabo cada consumidor.

Es importante señalar que, al igual que en la etapa de recuperación de datos, para llevar a cabo la generación, recuperación y aplicación de modelos de usuario también es necesario obtener el consentimiento y/o la autorización del usuario. De esta forma, de nuevo, el usuario, o el marco de datos personales actuando en su nombre, es el que determina la ejecución de las etapas presentadas (4.2.4 Cadena de valor).

En las siguientes secciones se presentan las contribuciones de esta tesis doctoral a las etapas mencionadas y su validación. Primero, la sección 7.1 presenta un análisis de las características generales de los modelos de usuario que el marco de datos personales debe producir para cubrir las necesidades de los distintos consumidores. Seguidamente, la sección 7.2 describe los patrones que pueden aplicarse para generar dichos modelos, considerando las relaciones que se establecen con el resto de entidades participantes y los procesos que deben llevarse a cabo. La sección 7.3 incluye dos trabajos realizados en el contexto de esta investigación que validan las características principales de los patrones de generación descritos. Cada una de las validaciones presenta los detalles de generación, recuperación y aplicación de los modelos de usuario asociados. Por último, se incluye un resumen del capítulo y las contribuciones incluidas en el mismo en la sección 7.4.

7.1 Características de los modelos de usuario del marco de datos personales

Como se indicó en el estado del arte (3.1.2 Perfil y modelo de usuario), los modelos de usuario pueden diseñarse e implementarse conforme a cuatro perspectivas diferenciadas que definen modelos que van de simples a complejos, de corto a largo plazo, de estereotipos a altamente adaptativos y de estáticos a dinámicos. Típicamente estas características responden a las particularidades y a las aplicaciones o los usos considerados por un proveedor de servicios concreto.

El marco de datos personales debe generar modelos de usuario que cubran las necesidades de diferentes consumidores de modelos de usuarios y es importante destacar que éstas pueden ser muy heterogéneas. Dadas las consideraciones anteriores, se establece que los modelos de usuario generados deberán incluir las siguientes características dentro de las cuatro perspectivas de un modelo de usuario (ver Figura 7-1):

- **Distintos niveles de complejidad.** La complejidad de un modelo de usuario es una característica subjetiva que depende de las capacidades de la entidad que lo valore. Independientemente de esto, se considerará que el sistema debe proporcionar modelos de usuarios simples que contengan datos personales recuperados de fuentes de datos externas (lo cual puede valorarse como un proceso complejo para un consumidor determinado) o modelos complejos obtenidos del procesamiento y la vinculación de datos heterogéneos. De esta forma, se considerarán modelos de usuario simples los que incluyen datos personales de los usuarios proporcionados y/u

observados en distintas fuentes de datos externas y modelos de usuario complejos los que incluyen datos derivados y/o inferidos por el marco de datos personales.

- **Valor a corto y largo plazo.** El sistema debe ser capaz de generar modelos de usuario que contengan información con distinta validez temporal y por lo tanto, los mecanismos de generación, así como de recuperación de los recursos, deben estar alineados con dicha validez. Se considerarán modelos de usuario con validez a corto plazo aquellos que incluyen o se generan a partir de información asociada a datos personales temporales como, por ejemplo, la información de contexto del usuario, la localización, las transacciones que está realizando, la actividad física y características biológicas asociadas en un momento determinado, entre otros. Por el contrario, los modelos de usuario con validez a largo plazo reflejan características persistentes y/o permanentes de los usuarios como las preferencias, competencias, roles, etc.
- **Altamente adaptativos.** Dado que el marco de datos personales cuenta, en principio, con los recursos necesarios para modelar a un usuario específico, se descarta el establecimiento de estereotipos que representen grupos de usuarios generales. Es decir, los modelos de usuario producidos por el marco de datos personales describen las particularidades y características de los usuarios en función de sus datos personales u otros recursos que permiten deducir o inferir dicha información y no es necesario la aplicación de estereotipos. Además, la definición de éstos supone un gran esfuerzo y el análisis de las características globales o los posibles tipos de usuarios finales de un sistema concreto, lo cual se distancia del objeto del marco de datos personales.

La generación de modelos de usuario altamente adaptativos requiere la utilización de los datos personales que refieren directamente con un usuario y esto se logra gracias al establecimiento de una infraestructura de gestión de identidad adecuada, tal y como se detallará en el capítulo 8. De esta forma, en este capítulo se asumirá que siempre se utilizan los datos relacionados con el usuario sobre el que se desea generar un modelo de usuario concreto.

- **Dinámicos.** Se considerarán únicamente modelos de usuario dinámicos que se actualizan a medida que lo hacen los datos personales asociados a un usuario. Es decir, los modelos de usuario variarán con el tiempo conforme lo hagan los datos personales de los usuarios y por ello, deberán utilizarse datos personales actualizados que reflejen las últimas características de los usuarios. Esta última característica se alcanzará gracias al establecimiento de mecanismos de recuperación adecuados que permitan obtener los recursos actualizados de los usuarios.

La siguiente figura resume lo expuesto en los puntos anteriores, señalando las características consideradas para los modelos de usuarios del marco de datos personales y las distintas categorías de datos personales asociados a éstas.

Figura 7-1. Modelos de usuario considerados en el marco de datos personales.

La consecución de dichas características se puede alcanzar a través de distintos patrones de generación de modelos de usuario que se describen en la siguiente sección.

7.2 Patrones para la generación de modelos de usuario

La generación de modelos de usuario engloba todos los procesos que deben llevarse a cabo para producir modelos de usuario. Dichos procesos están relacionados con el momento en el que se lleva a cabo la recuperación de los recursos necesarios para producir modelos de usuarios y con las características específicas de los modelos requeridos por los consumidores de modelos. Conforme a lo expuesto en la sección anterior (ver 7.1), las características que repercuten directamente en el establecimiento de los patrones de generación son principalmente las relacionadas con la complejidad y temporalidad de los modelos, puesto que para producir modelos dinámicos se debe establecer mecanismos de recuperación que permitan obtener datos actualizados de los usuarios y por otra parte, la producción de modelos de usuario altamente adaptativos se alcanza a través de modelos de gestión de identidad que permitan identificar a los usuarios tal y como se describirá en el siguiente capítulo.

Dadas las consideraciones anteriores y a diferencia del proceso de modelado de usuario tradicional, la etapa de generación de modelos de usuario deberá considerar las siguientes interrelaciones:

1. **Recuperación de los recursos necesarios.** La etapa de recuperación de datos descrita en el capítulo anterior no tiene por qué estar asociada al almacenamiento de los recursos identificados, aunque también puede contemplarse esta posibilidad. De esta forma, los recursos necesarios para la generación de modelos de usuario pueden encontrarse custodiados por fuentes externas o bien almacenados en una entidad central del sistema, puesto que previamente han sido recuperados. En definitiva, la generación de modelos puede conllevar que se recuperen los recursos necesarios dependiendo de dónde se encuentren éstos.

2. **Temporalidad de los modelos de usuario solicitados.** El marco de datos personales se comporta como una fuente de datos (y en particular, de modelos de usuario) externa a los consumidores de modelos y en este sentido, se establecen dos mecanismos principales para la recuperación de los modelos solicitados: por petición o por suscripción. Estos mecanismos repercutirán en la etapa de generación de modelos de usuario, puesto que ésta podrá inicializarse bien bajo la petición de un consumidor de modelos de usuario o bien porque se ha recuperado un dato personal que cubre o da pie a producir o actualizar modelos de usuario que satisfacen las necesidades de determinados consumidores que están suscritos a dichos modelos. En términos generales, en el primer caso, los modelos de usuario generados tendrán asociadas características con validez a largo plazo, mientras que en el segundo supuesto, los modelos de usuario cubren necesidades relacionadas con modelos de usuario de corto plazo.
3. **Complejidad de los modelos de usuario solicitados.** Tal y como se ha establecido anteriormente, el marco de datos personales deberá ofrecer modelos de usuario simples y complejos dependiendo de las necesidades de los consumidores. En el segundo caso, será necesario incluir procesos relacionados con la generación de conocimiento que permitan obtener nuevos datos derivados y/o inferidos a partir de los recursos recuperados.

De acuerdo con lo expuesto, en el contexto de esta investigación se han establecido tres patrones principales que permiten llevar a cabo la generación de modelos de usuario: centralizado, distribuido y desacoplado. Además, pueden existir otros patrones híbridos que adopten y combinen las particularidades de los anteriores. A continuación, se describen los patrones de generación mencionados indicando para cada uno de ellos las entidades participantes (módulos de la arquitectura y entidades externas), su idoneidad para cubrir las características establecidas para los modelos de usuario y los procesos asociados a la generación de modelos de usuario. Para concluir esta sección se incluye un resumen de los patrones identificados y su correspondencia con las características establecidas para los modelos de usuario.

7.2.1 Patrón centralizado

Para la definición del patrón centralizado se parte del supuesto de que se han establecido acuerdos entre distintos consumidores de modelos de usuario y el marco de datos personales para llevar a cabo la generación de modelos de usuario. De esta forma, la generación de modelos se realiza bajo la petición de los consumidores y la recuperación de los recursos necesarios se puede realizar previamente a la etapa de generación. Es decir, la etapa de generación de modelos en este patrón no está asociada con procesos de recuperación de datos, aunque sea necesaria, y la recuperación de los modelos de usuario se realiza a través de mecanismos de petición.

El patrón centralizado recibe este nombre debido a que los recursos necesarios para la generación de modelos de usuario, así como los propios modelos de usuario producidos, se encuentran almacenados en una entidad central gestionada por el marco de datos personales,

en concreto, el módulo Almacén de recursos personales (5.2.5 Almacén de recursos personales).

La Figura 7-2 refleja las entidades participantes en el patrón centralizado para la generación de modelos de usuario: Almacén de recursos y Generador. Asimismo, se indica la secuencia de los procesos que permiten llevar a cabo la generación de modelos de usuario y finalmente, su recuperación.

Figura 7-2. Patrón centralizado para la generación de modelos de usuario.

La secuencia 0 indica que previamente a la generación de modelos de usuario se ha llevado a cabo una etapa de recuperación de recursos que permite poblar el módulo Almacén de recursos. En concreto, el módulo Recolector es el encargado de proveer los datos personales y los recursos necesarios de distintas fuentes externas y adecuarlos para su almacenamiento en dicho módulo. Los mecanismos utilizados por el módulo Recolector podrán ser cualquiera de los indicados en 6.3 Recuperación de recursos en el marco de datos personales.

Dado que el módulo Almacén de recursos contiene una réplica de los datos custodiados por fuentes externas y que se establecen características dinámicas de los modelos de usuarios generados, la etapa de recuperación de datos debe entenderse como un proceso paralelo a la generación de modelos que permite tener actualizados los datos almacenados en el sistema. Es decir, aunque sobre la Figura 7-2 se ha representado como un único proceso, la recuperación de datos tiene asociada una gran complejidad y se extiende a lo largo de todo el proceso de modelado de usuario para poder mantener los recursos de los usuarios con las últimas características generadas.

Una vez que se dispone de los recursos necesarios en el Almacén de recursos, el módulo Generador, que estará compuesto de distintos componentes o generadores específicos, utiliza dichos recursos (secuencia 1 en la Figura 7-2) para producir modelos de usuario (secuencia 2) que dirige al Almacén de recursos personales (secuencia 3). Finalmente, los consumidores de modelos de usuario podrán recuperar los recursos incluidos directamente sobre el Almacén de recursos a través del módulo Recuperador (secuencia 4).

La centralización de los recursos recuperados por el marco de datos personales permite establecer una representación común para recursos heterogéneos y además, tiene grandes beneficios relacionados con la disponibilidad y el acceso a éstos. De esta forma, esto facilita la producción y recuperación de modelos de usuario complejos compuestos por datos derivados y/o inferidos, puesto que se pueden interpretar de forma consistente los recursos recuperados, así como facilitar el intercambio de los nuevos recursos obtenidos.

Dadas las consideraciones anteriores, los procesos asociados al patrón centralizado son principalmente el establecimiento de un estándar común que propicie el análisis y el procesamiento de los datos recuperados para producir otros nuevos en forma de modelos de usuario y además, los procesos relacionados con la generación de nuevo conocimiento. En este sentido, en 7.3.1 Patrón centralizado para generación de modelos de usuario socio-económicos se presenta una solución centralizada basada en tecnologías semánticas que facilita la interoperabilidad semántica y sintáctica de los recursos recuperados y la producción de modelos de usuario complejos.

Por último, en referencia a la validez temporal de los modelos de usuario generados a través del patrón centralizado, dado que se considera que éste se basa en la recuperación de los modelos a petición de los consumidores, será óptimo para modelos de largo plazo que contengan datos persistentes y/o permanentes de los usuarios. De otra forma, los consumidores de modelos pueden obtener modelos de usuario que no tengan validez en el momento de su recuperación y aplicación.

7.2.2 Patrón distribuido

El patrón distribuido para la generación de modelos de usuario se diferencia del patrón centralizado en que los recursos necesarios para la producción de modelos de usuario se mantienen en las fuentes de datos externas y por ello, la etapa de generación conlleva que se deban recuperar dichos recursos. Este patrón se ha establecido considerando que la etapa de generación de modelos comienza cuando un consumidor solicita un modelo de usuario que no ha sido generado previamente en el sistema a través de mecanismos de petición.

La Figura 7-3 representa gráficamente las entidades participantes en el patrón distribuido: Generador, Registro y Recolector, así como la secuencia de los procesos que permiten llevar a cabo la generación de modelos.

Figura 7-3. Patrón distribuido para la generación de modelos de usuario.

Tal y como se desprende de la figura, la etapa de generación comienza cuando el módulo Generador recibe a través del módulo Recuperador la petición de un consumidor de modelos de usuario (secuencia 1). Esta petición implica que el módulo Generador obtenga información en el módulo Registro sobre los recursos necesarios (secuencias 2 y 3) y solicite al módulo Recolector dichos recursos (secuencia 4). Es importante destacar en este punto, que la información que se obtiene del módulo Registro (secuencias 2 y 3) incluye los permisos necesarios para que el módulo Recolector realice la recuperación de recursos. De esta forma, el módulo Recolector puede obtener los recursos solicitados por el módulo Generador (secuencias 5 y 6).

Una vez que el módulo Generador recibe los recursos necesarios para la producción de los modelos puede llevar a cabo distintos procesamientos (secuencia 7) para producir el modelo de usuario solicitado y responder al consumidor de modelos con éste (secuencia 8).

En este enfoque, puesto que los datos necesarios para la generación de modelos de usuario se obtienen directamente de las fuentes de datos externas y se presupone que éstos incluyen las últimas características y datos personales de los usuarios, se garantiza que los modelos de usuario producidos son dinámicos. Es decir, siempre se utilizan recursos actualizados sobre los usuarios que mantienen las fuentes de datos externas y por lo tanto, se producen modelos de usuario dinámicos sin una gran complejidad.

No obstante, el patrón distribuido tiene una clara dependencia con las fuentes de datos externas, puesto que es necesaria la interacción con éstas para producir los modelos de usuario solicitados, y esto implica que puedan producirse latencias en las respuestas a los consumidores de modelos. A diferencia del patrón anterior, la etapa de generación implica que se lleve a cabo la recuperación de los recursos y la disponibilidad y acceso a dichos recursos depende directamente de las fuentes externas. Además, deben utilizarse mecanismos de mediación que permitan traducir los lenguajes y formatos utilizados por las fuentes origen que pueden aumentar el tiempo de respuesta del sistema. En este sentido, el módulo Registro facilita la implementación de dichas funcionalidades.

Por consiguiente, los procesos asociados al patrón distribuido se basan fundamentalmente en la recuperación de los recursos necesarios y la coordinación y traducción de éstos a los formatos establecidos por el sistema para responder correctamente a los consumidores. Dado que estos procesos añaden retardos en las respuestas a los consumidores de modelos de usuario, los procesamiento y mecanismos de generación de nuevo conocimiento deben ser lo más sencillos y rápidos posibles y por ello, se establece que el patrón distribuido es idóneo para modelos simples y/o complejos con datos derivados que no requieran procesamiento complejos con un gran número de operaciones.

Al igual que en el patrón centralizado, el patrón distribuido se basa en la recuperación de modelos de usuario bajo la petición de determinados consumidores y por ello, será óptimo para modelos a largo plazo con datos persistentes y/o permanentes de los usuarios que responden a las necesidades de los consumidores de modelos de usuario en el momento de su solicitud.

7.2.3 Patrón desacoplado

El patrón desacoplado se fundamenta en la generación de modelos de usuario a medida que se recuperan los recursos necesarios para producir modelos solicitados por consumidores que previamente han configurado distintos mecanismos de suscripción provistos por el marco de datos personales. Este patrón rompe con los enfoques anteriores basados en mecanismos de recuperación de modelos de usuario basados en petición. Además, los mecanismos de recuperación de los recursos necesarios se basan principalmente en mecanismos de suscripción que inician las fuentes de datos externas cuando detectan la generación de un evento o dato personal de relevancia.

La Figura 7-4 representa gráficamente el patrón desacoplado para la generación de modelos de usuario. Aquí, se observan los módulos del marco de datos personales participantes: Recolector y Generador, y la secuencia de los procesos que definen el patrón.

Figura 7-4. Patrón desacoplado para la generación de modelos de usuario.

A diferencia de los patrones descritos anteriormente, en este patrón, la etapa de generación comienza cuando una fuente de datos externa al marco de datos personales detecta la creación de un recurso al que el sistema está suscrito. De esta forma, la fuente de datos notifica al módulo Recolector de la existencia de un nuevo dato (secuencia 1). Tal y como se describió 6.4.1 Recuperación de datos personales en servicios de mensajería y redes sociales: el servicio de pago Posdata, la notificación de un evento puede implicar que se lleven a cabo otros procesos basados en peticiones para la recuperación del recurso, aunque éstos se han omitido sobre la figura por no ser el mecanismo habitual.

Una vez que se ha obtenido el recurso detectado en la fuente de datos, el módulo Recolector redirige el evento recibido al módulo Generador (secuencia 2) y éste puede llevar a cabo distintos procesamientos (secuencia 3) para producir un modelo de usuario que previamente ha sido solicitado a través de mecanismos de suscripción por un consumidor de modelos. Finalmente, el modelo de usuario producido se entrega a dicho consumidor a través del módulo Recuperador del sistema (secuencia 4).

Como puede observarse de la descripción anterior, el patrón desacoplado se apoya en la utilización de mecanismos de recuperación de recursos basados en suscripción tanto para los datos necesarios como para los modelos generados. Esto implica que previamente a la generación se activen y configuren dichos mecanismos. Así, un consumidor de modelos de usuario debería indicar su interés en un determinado modelo de usuario a través de servicios de suscripción provistos por el marco de datos personales y de igual forma, el marco de datos personales deberá configurar la suscripción a recursos externos de las fuentes de datos.

Puesto que los modelos de usuario generados a través de este patrón se basan en los recursos que se producen en tiempo real en una fuente de datos externa, los modelos producidos serán dinámicos, contando con las últimas características de los usuarios.

El patrón desacoplado permite la generación de modelos simples en un corto espacio de tiempo pero además, pueden llevarse a cabo distintos procesamientos complejos que permitan obtener datos derivados y/o inferidos sobre los datos personales recibidos e incluso sobre otros que pueden recuperarse de otras fuentes, lo cual se ha validado en el escenario presentado en 7.3.2. Al igual que en el caso anterior, la incorporación de procesamientos complejos aumentará el tiempo de respuesta a los consumidores de modelos de usuario.

Resulta evidente de las explicaciones anteriores que el patrón desacoplado permite la generación de modelos de usuario con validez a corto plazo, puesto que permite la generación de modelos de usuario a medida que se crean o generan los datos necesarios para su producción.

7.2.4 Patrones híbridos

Los patrones descritos anteriormente son generalizaciones que cubren los casos de generación de modelos de usuarios más habituales. Sin embargo, podrán considerarse otros patrones híbridos que adquieran parte de las características y procesos establecidos por los anteriores patrones.

A modo de ejemplo, se pueden considerar patrones que tomen como base los patrones distribuidos y/o desacoplados pero que también se apoyen en el módulo Almacén de recursos donde previamente se hayan almacenado recursos necesarios para la producción de modelos o donde se custodien modelos de usuarios generados.

En el contexto de esta investigación, se ha considerado un patrón híbrido que se apoya en la aplicación de los patrones distribuido y desacoplado para la generación de modelos de usuario. En 7.3.2 se presentan los detalles de validación relacionados.

7.2.5 Patrones de generación y características de los modelos de usuario asociados

En los apartados anteriores se han descrito tres patrones principales que permiten llevar a cabo la generación de modelos de usuario con distintas características de temporalidad y complejidad. Además, los patrones descritos permiten producir modelos de usuario altamente adaptativos (que representan las particularidades de un usuario concreto) dada la infraestructura de gestión de identidad que se presentará en el siguiente capítulo y éstos tendrán características dinámicas (basados en la utilización de valores actualizados de los datos personales de los usuarios) por la aplicación de distintos mecanismos de recuperación.

Dadas las consideraciones anteriores, en esta sección, en primer lugar, se presentó el **patrón centralizado** (ver 7.2.1) que principalmente se sustenta en una entidad central (módulo Almacén de recursos) para llevar a cabo la generación de modelos de usuario. Esta entidad facilita la generación de modelos de usuario complejos con datos derivados e inferidos que refieren a un usuario concreto. Además, el patrón es idóneo para la generación de modelos con características de largo plazo, puesto que se establece bajo mecanismos de petición de modelos de usuario. No obstante, dado que los datos utilizados en la etapa de generación son una réplica de los originados por los usuarios en la fuente de datos, producir modelos de usuario con características dinámicas a través de este patrón añade complejidad a la etapa de recuperación de recursos, puesto que deben implementarse mecanismos que permitan tener actualizados dichos datos conforme lo hacen en la fuente origen.

Por su parte, el **patrón distribuido** (ver 7.2.2) está enfocado a soluciones de generación de modelos de usuario que implican la recuperación de los recursos necesarios de fuentes de datos externas y esto repercute en el tiempo de respuesta a los consumidores de modelos de usuario. Por ello, los procesos asociados a este patrón están relacionados con la obtención de los datos de distintas fuentes y procesamientos simples que no añadan retardo en la respuesta a los consumidores. Es decir, este patrón permite obtener modelos de usuario simples o complejos con datos derivados que no requieran un gran número de operaciones de procesamiento. Al igual que en el caso anterior, el patrón distribuido se focaliza en modelos de largo plazo con datos persistentes y/o permanentes de los usuarios, puesto que los mecanismos de recuperación de modelos de usuario se basan en peticiones. En este patrón, a diferencia del centralizado, dado que se utiliza para la generación los datos recuperados directamente de fuentes externas, se garantiza características dinámicas para los modelos fácilmente.

Finalmente, el **patrón desacoplado** (ver 7.2.3) se sustenta en la utilización de mecanismos de suscripción que permiten, por una parte, recuperar los datos necesarios para la generación a medida que se producen en la fuente origen y así, por otra parte, entregar los modelos de usuario producidos sobre dichos datos a los consumidores. De esta forma, el patrón centralizado permite entregar modelos de usuario de corto plazo a los consumidores suscritos y éstos pueden ser simples y/o complejos dependiendo de los requisitos establecidos por los consumidores. Al igual que en el patrón distribuido, los modelos de usuario tienen características dinámicas, puesto que se utilizan los últimos datos asociados a un usuario.

La Tabla 7-1 resume lo expuesto anteriormente indicando para cada patrón establecido las características de los modelos de usuario asociados. Es importante destacar de nuevo que todos ellos permiten la producción de modelos de usuario altamente adaptativos, que refieren a un usuario concreto, dada la infraestructura de identidad establecida para el marco de datos personales (ver capítulo 8) y que todos los modelos tendrán características dinámicas, puesto que se presupone el establecimiento de mecanismos de recuperación adecuados (ref. cap. 6).

Tabla 7-1. Patrones de generación y características de los modelos de usuario asociados.

Patrón	Características modelos de usuario generados
Centralizado	<ul style="list-style-type: none"> • Modelos simples (datos proporcionados y observados) y complejos (datos derivados e inferidos) • Largo plazo (datos persistentes y permanentes) • Altamente adaptativos • Dinámicos (añade complejidad en la etapa de recuperación de datos)
Distribuido	<ul style="list-style-type: none"> • Modelos simples (datos proporcionados y observados) y complejos (datos derivados) • Largo plazo (datos persistentes y permanentes) • Altamente adaptativos • Dinámicos
Desacoplado	<ul style="list-style-type: none"> • Modelos simples (datos proporcionados y observados) y complejos (datos derivados e inferidos) • Corto plazo (datos temporales) • Altamente adaptativos • Dinámicos

A continuación, en la siguiente sección, se describen las validaciones realizadas de los patrones de generación presentados y que globalmente cubren las características establecidas para los modelos de usuario en cuanto a su complejidad (modelos simples y/o complejos) y temporalidad (modelos de corto y/o largo plazo), todos ellos con características dinámicas y altamente adaptativos. Principalmente, dichas validaciones permiten dar forma al modelo de negocio descrito en 4.3.2 Modelo de negocio para la promoción de nuevos servicios financieros.

7.3 Validaciones de los patrones de generación para la creación de nuevos servicios financieros

En esta sección se presentan las validaciones realizadas de los patrones de generación de modelos de usuario descritos en la sección anterior. En concreto, en el apartado 7.3.1 se incluye una validación del patrón centralizado basada en tecnologías semánticas que permite producir modelos socio-económicos para ofrecer nuevos servicios financieros. Seguidamente, en el apartado 7.3.2 se presenta un escenario de validación para un patrón híbrido que combina el patrón desacoplado y distribuido. Este patrón permite generar modelos de usuario enriquecidos de corto plazo para mejorar y ofrecer nuevos servicios financieros a una entidad bancaria. Las validaciones incluidas se basan en los recursos recuperados en 6.4 Validación de componentes específicos para la recuperación de datos y además en de otras fuentes de datos que pueden ser consideradas como se detallará a continuación.

7.3.1 Patrón centralizado para generación de modelos de usuario socio-económicos

En el contexto de esta investigación, se ha llevado a cabo una validación del patrón centralizado para la generación de modelos de usuario socio-económicos basada en el uso de tecnologías semánticas, cuyo objetivo es fomentar la creación de nuevos servicios financieros haciendo uso de los recursos ofrecidos por distintas fuentes de datos externas tal y como se describió en 4.3.2 Modelo de negocio para la promoción de nuevos servicios financieros.

Las tecnologías semánticas han aparecido para solventar problemas de interoperabilidad, puesto que facilitan la reutilización, agregación, correspondencia, fiabilidad, publicación, generación y la definición común de información (Uschold y Gruninger, 1996). En particular, dentro de éstas, las ontologías permiten la representación de conocimiento de forma unificada, simplificando la traducción de distintos vocabularios, y así, facilitan el intercambio de datos e información entre distintos sistemas.

En un patrón centralizado (7.2.1 Patrón centralizado), el establecimiento de un estándar común sobre las características de los usuarios que propicie y favorezca el análisis y el procesamiento de los datos recuperados para la generación de modelos de usuario constituye uno de los procesos más relevantes. En este sentido, el uso de ontologías permite expresar dicho estándar y así, establecer un vocabulario consensuado entre los distintos actores participantes, garantizando la interoperabilidad semántica y sintáctica en el sistema. Además, la información representada a través de ontologías puede ser procesada y utilizada para generar otra nueva dadas las relaciones expresadas entre los distintos conceptos. De esta forma, las ontologías permiten establecer un Modelo de usuario interoperable del que pueden derivarse y obtenerse otros nuevos modelos de usuario.

La Figura 7-5 representa el escenario de validación del patrón centralizado, incluyendo las etapas principales del proceso de modelado de usuario distribuido: recuperación de datos, generación y aplicación de modelos, así como los módulos principales de las entidades participantes. En particular, el marco de datos personales (en la figura PeDF) incluye los

conocimiento, en ocasiones tácito, existente en su red social próxima (lo que se ha dado en llamar *friendsourcing* (Bernstein, et al., 2008)).

El servicio FriendLoans incluye un módulo Recuperador que permite obtener y solicitar modelos de usuario socio-económicos al marco de datos personales y que aplica para proveer el servicio a sus usuarios. Además, FriendLoans incorpora un módulo Recolector que **realimenta el proceso de modelado** de usuario distribuido con los datos generados por sus usuarios. Es decir, el servicio FriendLoans actúa también como fuente de datos externa al marco de datos personales proporcionando nuevos datos que enriquecen el Modelo de usuario interoperable establecido.

En los siguientes apartados se detallan las etapas y componentes principales del escenario de validación presentado, destacando los trabajos relacionados con las últimas etapas del proceso de modelado de usuario distribuido, puesto que la recuperación de los datos necesarios ya se describió en el capítulo anterior. En primer lugar, se presenta la particularización del Modelo de usuario interoperable realizada para el escenario presentado. Seguidamente, se describen los mecanismos de generación asociados y el servicio FriendLoans que lleva a cabo la recuperación y aplicación de distintos modelos de usuario. Por último, se exponen los detalles de implementación relacionados con el escenario de validación.

7.3.1.1 Modelo de usuario interoperable

Dado el escenario de validación presentado, se ha establecido un Modelo de usuario interoperable basado en una ontología, denominada como *Personal Socio-Economic Network* (PSEN), que ha sido desarrollada en el contexto de esta investigación y que define las características demográficas y sociales de las personas y el intercambio de dinero que se produce entre éstas.

Para mejorar la interoperabilidad semántica y sintáctica de los datos representados, la ontología PSEN se basa en la utilización de otras ontologías existentes. En este sentido, para el contexto social de un usuario, existen distintas definiciones y representaciones pero una de las más relevantes y adoptadas es la ontología FOAF (Brickley y Miller, 2014) y por ello, se ha identificado como la mejor alternativa para su reutilización.

Por otra parte, respecto a las características financieras y actividades de económicas de los usuarios, se han identificado distintas propuestas como los estándares de servicios financieros del *Object Management Group* (OMG) y en concreto, la *Financial Industry Business Ontology* (FIBO) (OMG, 2015), otras relacionadas con productos específicos de detección de fraudes o recomendación de productos financieros como las incluidas en (Tang y Song, 2011), (Gonzalez-Carrasco, et al., 2012) o (Ying, et al., 2009) y distintos estándares financieros como el *eXtensible Business Reporting Language* (XBRL)⁸³ o el *Financial products Markup Language* (FpML)⁸⁴. Sin embargo, dichas representaciones están enfocadas a soluciones y servicios muy específicos de la industria financiera y se alejan de las características particulares del usuario.

⁸³ XBRL, The business reporting standard, <https://www.xbrl.org/>

⁸⁴ FpML, <http://www.fpml.org/>

Dadas las consideraciones anteriores, se ha optado por adoptar una ontología de alto nivel (Carrión Delgado, 2012) y en concreto, la ontología *Suggested Upper Merged Ontology* (SUMO) (Niles y Pease, 2001), que describe conceptos generales (por ejemplo, del espacio, tiempo, agentes, etc.) independientes, aunque comunes, de un problema particular o dominio. Además, la ontología SUMO tiene una estructura modular e incluye distintas subontologías y ontologías específicas de distintos dominios como el financiero. De esta forma, se ha utilizado la ontología financiera definida por SUMO y además, la nomenclatura utilizada en la definición de los conceptos de PSEN se basa en los términos incluidos en la ontología de alto nivel SUMO, lo que permite añadir nuevos conceptos y otras ontologías de distintos dominios no considerados en la validación fácilmente.

Básicamente, la ontología PSEN incluye un conjunto de términos y las relaciones entre éstos para describir a las personas, las relaciones sociales entre éstas, sus datos financieros y las actividades económicas o intercambios de dinero que se producen entre distintos usuarios (Figura 7-6).

Figura 7-6. Resumen ontología Personal Socio-Economic Network (San Miguel, et al., 2015a).

Cada una de las personas o usuarios del marco de datos personales se representa a través de la clase *Person* de la ontología FOAF que tiene asociada distintas propiedades para describir su información demográfica y de identificación como, por ejemplo, nombre (*firstName*), apellido (*lastName*), sexo (*gender*), edad (*age*), día y mes de nacimiento (*birthday*) y correo electrónico (*mbox*). Estas últimas características han sido omitidas de la figura para facilitar su legibilidad.

Dada su relevancia para el escenario de validación, lo que sí se ha representado en la figura es la propiedad *knows* de FOAF que indica la existencia de una relación entre dos personas. En FOAF, esta propiedad toma su sentido más amplio y no detalla el grado de las relaciones indicadas, aunque implica una relación recíproca entre dos personas. Existen distintos trabajos relacionados con este último punto y cuyo objetivo es extender la ontología FOAF para añadir

los grados de confianza, las relaciones y/o roles que se establecen entre dos personas. Ejemplo de esto último son los incluidos en (Golbeck, et al., 2003), (Zhang y Yu, 2012) o (Kabir, et al, 2014).

Además de los elementos anteriores, la clase *Person* tiene asociada distintas clases *Online Account* de la ontología FOAF que permite modelar la información de identificación y los servicios telemáticos que utiliza una persona específica. En la ontología PSEN, esta clase ha sido extendida para incluir otras dos nuevas:

- *Online Payment Account*. Representa servicios telemáticos que permiten a los usuarios realizar operaciones de pago a través de Internet como, por ejemplo, el servicio Posdata. En términos generales esta clase tiene asociado una tarjeta bancaria (*BankCard*) o una cuenta financiera (*Financial Account*) que denotan dónde se hace efectivo el pago realizado. Ambas clases están definidas dentro de la ontología SUMO y se relacionan a través de la propiedad *cardAccount*, puesto que una tarjeta bancaria habitualmente está asociada a una cuenta financiera.
- *Online Banking Account*. Representa servicios bancarios telemáticos ofrecidos por distintas instituciones financieras como el Banco Santander y tiene asociada una cuenta financiera (*Financial Account*).

Para modelar las actividades económicas de los usuarios, se ha tomado como partida los conceptos incluidos en la ontología SUMO y éstas se han considerado como interacciones sociales que ocurren entre dos personas. En concreto, se ha definido una clase denominada como *SocialInteraction* que tiene asociada a dos personas a través de las propiedades *hasOrigin* (indica la persona que inicializa la interacción) y *hasDestination* (la persona que forma parte de la interacción o el sujeto pasivo de ésta). Conjuntamente, la clase *SocialInteraction* incluye tres propiedades principales: fecha (*timestamp*), canal (*channel*) y participante (*patient*). Las dos primeras propiedades indican cuándo y dónde ocurre la interacción social respectivamente, mientras que la propiedad *patient* designa entidades o agentes (representada como la clase *Entity*) que participan en la interacción como, por ejemplo, la cantidad de dinero que intercambian dos personas.

La clase *SocialInteraction* incluye dos subclases que representan transacciones (*Transaction*) y comunicaciones (*Communication*) y que tienen respectivamente las subclases pagos (*Payment*) y solicitudes (*Request*) asociadas. Estas últimas se relacionan a través de la propiedad *hasPayment* para indicar que una solicitud de dinero ha sido pagada.

En conclusión, la ontología PSEN permite constituir redes económicas personales de distintos usuarios, representando no sólo sus características particulares y sociales, sino también las relaciones económicas que se establecen entre ellos.

7.3.1.2 Generación de modelos de usuario

Dado que el Modelo de usuario interoperable se basa en la utilización de distintas ontologías y en particular, en la ontología PSEN para el escenario presentado, pueden utilizarse distintos mecanismos que permiten recuperar los datos almacenados y obtener otros nuevos.

En cuanto a la recuperación de datos almacenados, existen distintos lenguajes como, por ejemplo, SPARQL, que permiten obtener los datos almacenados a través de ontologías y así, obtener **modelos de usuario simples**. La recuperación a través de este tipo de lenguajes generalmente permite añadir una serie de restricciones y obtener propiedades concretas de una instancia o una lista de resultados. En particular, las especificaciones relacionadas con SPARQL incluyen un lenguaje de consulta, un formato de las respuestas y un medio para el transporte de las peticiones y respuestas.

Por otra parte, uno de los aspectos más relevantes de las ontologías es que permiten razonar o derivar e inferir nueva información a partir de la disponible, lo cual puede utilizarse para generar **modelos de usuarios complejos**. Existen distintos métodos que permiten llevar a cabo dichas capacidades como son: el razonamiento con lógica de primer orden, razonamiento con lógica descriptiva y el razonamiento basado en reglas (Flores Vitelli, 2011).

El razonamiento a través de lógica de primer orden se fundamenta en la representación de objetos de un dominio, las relaciones entre éstos y el establecimiento de distintos predicados para realizar inferencias sobre dicho dominio. De esta forma, se determinan nuevos hechos como consecuencia lógica de los ya existentes.

Por su parte, la lógica descriptiva es un subconjunto de la lógica de primer orden que se apoya en distintos lenguajes para describir conocimiento y uno de sus objetivos principales es comprobar que una base de conocimiento no incluye proposiciones contradictorias.

En último lugar, el razonamiento basado en reglas, como su propio nombre indica, se apoya en utilización de distintas reglas que representan los elementos de razonamiento y que son establecidos por expertos del dominio. De esta forma, a través de este tipo de razonamiento se infiere nueva información tal y como lo haría un experto del dominio.

Básicamente, los sistemas que aplican el razonamiento basado en reglas cuentan con una base de hechos, una base de reglas y un motor de inferencia (Figura 7-7). La base de hechos engloba el conocimiento sobre un dominio y se va modificando a medida que el motor de inferencia aplica las distintas reglas establecidas. Dependiendo del problema que se quiera solventar o de los resultados deseados, se pueden aplicar distintas estrategias que determinan el orden y parada de ejecución de las reglas y los mecanismos de inferencia que pueden aplicarse. Principalmente, existen dos mecanismos de inferencia: encadenamiento hacia delante (cuando se enfrentan las reglas a la base de hechos) o encadenamiento hacia atrás (se parte de la conclusión a la que se quiere llegar y se busca la regla apropiada).

Figura 7-7. Elementos para el razonamiento basado en reglas (Flores Vitelli, 2011).

El módulo Generador del marco de datos personales puede englobar distintos razonadores semánticos como los descritos anteriormente que permiten producir nuevos datos y que una vez generados se almacenan en el Modelo de usuario interoperable. En este sentido, el establecimiento de los distintos razonadores depende de las necesidades específicas de los consumidores de modelos de usuario.

Para el caso concreto del servicio FriendLoans se ha utilizado un razonador basado en reglas y éstas se han definido de acuerdo con las necesidades indicadas por el servicio. Por ejemplo, algunas de las reglas incluidas permiten detectar si un usuario debe dinero a otro, si un usuario ha recibido una cantidad mayor de X euros y si un usuario ha solicitado una cantidad de dinero superior a Y euros. En las reglas, los valores A , X e Y son parámetros que los consumidores de modelos de usuario pueden personalizar en su consulta.

7.3.1.3 Aplicación de modelos de usuario: el servicio FriendLoans

Las últimas etapas del proceso de modelado de usuario distribuido (recuperación y aplicación de modelos) se llevan a cabo a través de un nuevo servicio financiero desarrollado en el contexto de esta investigación, el servicio FriendLoans.

FriendLoans es un servicio basado en el concepto de economía colaborativa mediante micromecenazgo o *crowdfunding* (3.6.2 Modelos de negocio en el contexto de servicios telemáticos) y en concreto, en el denominado *friendsourcing*, que ofrece a sus usuarios la posibilidad de recibir a través de un conjunto de micropréstamos de sus amigos una determinada cantidad de dinero. Para ello, el servicio recomienda aquellos posibles prestadores que satisfacen un grado de confianza, liquidez y solvencia dados.

El funcionamiento del servicio consiste, en primer lugar, en que los usuarios registrados indican desde su página principal la cantidad que necesitan (parte superior de la Figura 7-8) y seguidamente, el servicio FriendLoans responde con una lista de amigos que considera óptimos como prestamistas (parte inferior de la Figura 7-8). Adicionalmente, los usuarios pueden indicar para qué necesitan el dinero, cuándo podrán devolverlo y otros comentarios y condiciones añadidas.

Figura 7-8. Captura de pantalla del servicio FriendLoans (San Miguel, et al., 2015c).

Sobre la lista de posibles prestamistas, los usuarios pueden solicitar individualmente a cada uno de sus amigos la aportación de un micropréstamo. Cada una de estas solicitudes tiene asociada distintas condiciones que establece el prestamista y que puede negociar con el usuario que solicita el préstamo. Por ejemplo, se pueden añadir contraprestaciones como intereses de crédito o vincular la devolución del micropréstamo a una fecha concreta.

El pago final del micropréstamo se puede realizar a través de distintos canales tradicionales como, por ejemplo, transferencias, u otros servicios de pago telemáticos personales como el servicio Posdata.

En definitiva, el servicio FriendLoans debe incorporar los mecanismos necesarios para permitir a un usuario solicitar una cantidad de dinero determinada, obtener una lista de posibles prestamistas, formalizar las negociaciones de las condiciones de cada micropréstamo y finalmente, ejecutar los pagos entre usuarios.

La generación de la lista de posibles usuarios implica que el servicio conozca características sociales y situaciones financieras individuales de cada uno de sus usuarios y para ello, recurre a la información almacenada y generada en el marco de datos personales y en particular, en el Modelo de usuario interoperable, a través del módulo Recuperador (ver Figura 7-5).

El módulo Recuperador se comunica con el marco de datos personal a través de distintas interfaces que abstraen de los detalles concretos relacionados con las etapas de recuperación de datos y generación. Así, este módulo obtiene modelos de usuario realizando distintas llamadas al Modelo de usuario interoperable del marco de datos personales. Estas llamadas se pueden personalizar, variando distintos parámetros, para obtener modelos de usuario a

medida que pueden ser simples o complejos. Por ejemplo, se puede solicitar la lista de amigos de un usuario específico o llamadas más complejas donde se solicite la lista de amigos que son buenos prestadores para una cantidad determinada.

En cuanto al resto de módulos del servicio FriendLoans, el módulo Administrador de usuarios permite que los usuarios lleven a cabo las funcionalidades principales del servicio (solicitud de dinero, petición de los micropréstamos o pago de éstos, entre otras) y se comunica con el módulo Recuperador para mostrarle al usuario solicitante los posibles prestamistas. El módulo Gestor de Solicitudes recoge las peticiones de dinero de un usuario a sus amigos y se comunica con el módulo Gestor de Pagos para que estos últimos realicen los pagos a través de distintos servicios financieros cuando existe un acuerdo en la condiciones del préstamo. Además, el módulo Gestor de Solicitudes redirige la información relacionada con cada una de las solicitudes de micropréstamos efectuadas por un usuario al módulo Recolector que envía dicha información al módulo Recolector del marco de datos personales para realimentar y enriquecer el Modelo de usuario interoperable con nuevos datos personales.

7.3.1.4 Detalles de implementación

El escenario de validación presentado se ha llevado a cabo a través de distintas herramientas de código abierto y desarrollos propios que han permitido implementar los componentes principales del marco de datos personales para el patrón centralizado, así como del servicio FriendLoans.

El marco de datos personales se apoya en el entorno de desarrollo de aplicaciones semánticas Jena (3.2.2.4 Modelos basados en tecnologías semánticas) y sus distintas bibliotecas. En particular, el almacenamiento semántico del Modelo de usuario interoperable se ha realizado con el componente de *Jena TDB*, siguiendo la definición dada por la ontología PSEN y con una estructura de grafos RDF. Para la definición de la ontología PSEN, se ha utilizado la herramienta Protégé que permite la definición de ésta a través del lenguaje OWL.

Por otra parte, el marco de datos personales utiliza la *Inference API* de Jena para implementar los distintos componentes del módulo Generador que obtiene nuevo conocimiento a partir de la información previamente almacenada en el Modelo de usuario interoperable. Jena incluye un conjunto de razonadores predefinidos, un razonador genérico basado en reglas y los mecanismos necesarios para incluir razonadores externos. En el caso concreto del escenario de validación, se ha configurado el razonador genérico de Jena y se han implementado distintas reglas de inferencia. Además, se han realizado otros procesamientos experimentales con el razonador externo Pellet.

Las reglas de inferencia implementadas en el marco de datos personales siguen una sintaxis similar a las del sistema experto *C Language Integrated Production System (CLIPS)*⁸⁵ y son de la forma “Si A entonces B”. Es decir, si se cumple una serie de condiciones A entonces se concluye B.

⁸⁵ CLIPS, <http://clipsrules.sourceforge.net/>

A modo de ejemplo, se incluye una de las reglas implementadas que determina si los amigos de un usuario determinado (en el ejemplo, Alice) han recibido un pago mayor de una cantidad (en el ejemplo, 10€) y por lo tanto, si éstos tienen liquidez para realizar un micropréstamo. Esta regla se ha llamado *hasPaymentBiggerThan* y sus condiciones se expresan en las líneas 2ª-8ª, éstas son: el usuario Alice debe conocer a la persona, esa persona debe tener pagos mayores de 10€. El resultado es la lista de personas que cumplen con dichas condiciones (indicado en la línea 10ª).

```

1: [hasPaymentBiggerThan:
2: (?Alice psen:isTarget "true"^^xs:boolean)
3: (?Alice psen:knows ?person)
4: (?socInt psen:hasDestination ?person)
5: (?socInt sd: psen:Payment)
6: (?socInt psen:patient ?entity)
7: (?entity psen:amount ?amount)
8: greaterThan(?amount,10)
9: ->
10: (?person psen:hasPaymentBiggerThan "true"^^xs:boolean)]

```

En Jena, las reglas no siguen una jerarquización establecida y por consiguiente, no tienen un orden de ejecución predecible, lo que permite añadir nuevas reglas fácilmente.

Finalmente, se ha utilizado el servidor SPARQL de Jena, *Apache Jena Fuseki*, para ofrecer a terceras entidades la información del Modelo de usuario interoperable del marco de datos personales. Fuseki soporta peticiones SPARQL a través del protocolo de comunicación HTTP. Cada consumidor de modelos de usuario debe implementar las llamadas específicas al Modelo de usuario interoperable. En el caso del servicio FriendLoans, se ha implementado una llamada que solicita para un usuario determinado la lista de sus amigos, priorizando esta lista con los amigos que tienen deudas con el usuario al que aplica el modelo y que tienen suficiente liquidez (han recibido un pago mayor de una cantidad) para realizar el micropréstamo.

7.3.2 Patrón desacoplado y distribuido para la generación de modelos de usuario enriquecidos de corto plazo

En esta sección se presenta la validación realizada para el patrón desacoplado y distribuido de generación de modelos de usuario que permite obtener modelos enriquecidos de corto plazo. En concreto, los modelos generados se basan en los eventos personales de negocio que se producen dentro de una arquitectura empresarial del ámbito bancario (6.4.2.1 Descripción de la arquitectura empresarial BankSphere), su recuperación y la combinación con otros datos para producir modelos de usuario de corto plazo que pueden ser utilizados por distintos consumidores.

Tal y como se describió en 7.2.3, el patrón desacoplado se basa en la utilización de mecanismos de suscripción para la recuperación de los datos necesarios en la generación y para la entrega de los modelos de usuario producidos, lo que permite generar modelos de corto plazo (lo que es lo mismo, formados por datos que tienen especial relevancia en un determinado momento). Además, estos modelos pueden ser simples o complejos dependiendo de los procesos realizados por el módulo Generador (secuencia 3 en Figura 7-4). En el

contexto de esta validación, se ha considerado que dicho módulo se apoya en la recuperación de distintos datos conforme a lo establecido por el patrón distribuido (ver 7.2.2), lo cual permite recuperar datos de fuentes externas que pueden ser utilizados para el enriquecimiento de los datos recuperados previamente y con ello, producir modelos derivados de la combinación de todos los datos obtenidos.

La Figura 7-9 representa el escenario propuesto para la validación del patrón desacoplado y distribuido, indicando la secuencia que permite producir modelos de usuario enriquecidos de corto plazo. Aquí, se han representado las tres etapas del proceso de modelado de usuario distribuido (recuperación, generación y aplicación) y se distingue la participación de distintas fuentes de datos externas (Productor 1, Productor 2 y otras fuentes que contienen recursos para el enriquecimiento), el marco de datos personales (representado como *PeDF* en la figura) y distintos consumidores de modelos (una herramienta de visualización de datos, un *Customer Relationship Management* – CRM y un servicio de alertas del usuario).

Figura 7-9. Escenario de validación para los patrones desacoplado y distribuido de generación de modelos los de usuario.

En el escenario de validación representado, el proceso de generación de eventos se inicia cuando un usuario del marco de datos personales utiliza distintos servicios ofrecidos por una entidad bancaria (secuencia 1 en la figura) y esto provoca que se genere un evento personal de negocio en el servicio utilizado tal y como se indicó en el capítulo 6. De esta forma, el marco de datos personales a través del módulo Recolector recibe el evento generado por el usuario (secuencia 2) y éste lo redirige al módulo Generador (secuencia 3) del marco de datos personales que será el encargado de producir modelos de usuario enriquecidos para los distintos consumidores de eventos interesados.

Se ha considerado que para la generación de modelos de usuario a partir de los eventos personales de negocio recuperados, el módulo Generador requiere otros datos que permitan su enriquecimiento. Para ello, solicita al módulo Recolector los datos necesarios (secuencia 4), éste los recupera de distintas fuentes externas que los custodian (secuencia 5) y los entrega al módulo Generador (secuencia 6). Con todos estos datos a su disposición, el módulo Generador puede llevar a cabo una serie de procesamientos (secuencia 7) que permite producir modelos de usuario que son notificados a los consumidores suscritos a éstos (secuencia 8).

En el escenario de validación se han considerado dos tipos de eventos personales de negocio: envío y recepción de transferencias para un usuario. A continuación, se describen los mecanismos de generación que permiten producir modelos de usuario a través de este patrón y los detalles de aplicación o uso de éstos por parte de distintos consumidores de modelos de usuario.

7.3.2.1 Generación de modelos de usuario

En el escenario presentado, la generación de modelos de usuario puede llevarse a cabo mediante la utilización de distintos mecanismos. En particular, se ha considerado la incorporación de motores de procesamiento de eventos y el establecimiento de un patrón distribuido para llevar a cabo el enriquecimiento de los eventos personales de negocio recuperados.

Los **motores de procesamiento de eventos** (ver 3.3.4.1 Arquitectura dirigida a eventos) permiten realizar un conjunto de operaciones sobre los eventos recibidos como, por ejemplo, la lectura, creación, transformación, correlación, etc. Sus aplicaciones son muy diversas y principalmente, se pueden agrupar dentro de cinco categorías principales (Etzion y Niblett, 2010):

- observación para detectar comportamientos anómalos a partir de los eventos recuperados,
- difusión de información conforme a requisitos previos establecidos,
- reacción para llevar a cabo determinadas acciones,
- diagnóstico a partir de un supuesto establecido, y
- predicción de otros eventos antes de que ocurran.

La configuración de un motor de procesamiento puede realizarse a través de distintos lenguajes como, por ejemplo, aquellos basados en reglas, orientados al flujo de eventos o imperativos.

En el contexto de este trabajo, se ha utilizado la herramienta de procesamiento de eventos Esper⁸⁶ que se distribuye mediante licencia GPL y que además cuenta con una versión comercial. Ésta se basa en un lenguaje orientado al flujo de eventos para realizar procesamiento en *stream* y procesamiento complejo de eventos. En particular, Esper se apoya

⁸⁶ EsperTech, <http://www.espertech.com/>

en el lenguaje *Event Query Language* (EQL) que tiene una sintaxis similar al SQL de bases de datos. EQL utiliza el flujo de eventos y distintas ventanas temporales como fuente de datos de búsqueda y esto permite realizar el filtrado, agregación y unión en un flujo de eventos. Además, la herramienta permite detectar distintos comportamientos en el flujo de eventos basándose en patrones y correlar información de distintos flujos.

En particular, se ha utilizado Esper para obtener información de dos tipos de eventos personales de negocio: envío y recepción de transferencias. Por ejemplo, se ha obtenido el número de transferencias enviadas y recibidas por segundo, el importe acumulado de éstas, la localización de las entidades bancarias desde donde se realizan las transferencias, las cinco con mayor número de transferencias asociadas, etc. Además, Esper permite reaccionar ante determinados eventos y en este sentido, se ha utilizado para poder llevar a cabo la recuperación de datos custodiados en otras fuentes. En relación a esto, se ha consultado una base de datos de clientes que permite obtener información sobre el perfil de un determinado cliente de la entidad y determinar si tiene activado un servicio de alertas, el canal seleccionado como preferido para sus relaciones con la entidad, etc.

7.3.2.2 Aplicación de modelos de usuario

Tal y como se reflejó en la representación del escenario de validación (ver Figura 7-9), se han considerado tres consumidores de modelos de usuario que aprovechan y reutilizan los datos generados por el marco de datos personales a partir de los eventos personales de negocio recuperados. Estos consumidores son: una herramienta de visualización de datos, un sistema de gestión de relaciones con los clientes (del inglés, *Customer Relationship Management* – CRM) y un servicio de alertas del usuario. A continuación se describen brevemente dichos sistemas y el uso que realizan de los modelos generados por el marco de datos personales.

Visualización de datos

Con el objeto de monitorizar la actividad de negocio de la arquitectura empresarial BKS presentada en el capítulo anterior, se ha implementado una herramienta de visualización de datos que recibe los eventos personales de negocio e información obtenida a partir de éstos del marco de datos personales y los representa gráficamente.

La siguiente figura muestra una captura de pantalla de la herramienta de visualización.

Figura 7-10. Herramienta de visualización para el consumo de eventos personales de negocio e información relacionada.

La implementación de la herramienta se ha realizado mediante la tecnología *WebSocket* (W3C, 2012) que permite obtener la información generada en el marco de datos personales a medida que se produce. Además, se ha utilizado *Google Charts*⁸⁷ para representar las distintas gráficas, tablas y mapas incluidos.

Sistema de gestión de relaciones con los clientes

Además del consumidor anterior, se ha considerado un sistema de gestión de relaciones con los clientes o CRM que recibe información relevante sobre un usuario específico del marco de datos personales para tomar determinadas acciones comerciales.

En particular, el CRM recibe un evento personal de negocio enriquecido con datos de otras fuentes externas de la arquitectura empresarial BKS que custodian información relacionada con el usuario. De esta forma, por ejemplo, puede recibir información que indique si un determinado usuario ha hecho una transferencia mayor de una determinada cantidad y el número de teléfono del usuario para que el operador del CRM realice distintas acciones comerciales.

Servicio de alertas al usuario

También se ha considerado como consumidor de modelos de usuario de este escenario un servicio de alertas que está suscrito al mismo tipo de eventos que en el caso presentado anteriormente y que permite emitir distintas notificaciones al usuario dependiendo de su perfil.

⁸⁷ Google Developers, Charts, <https://developers.google.com/chart/>

Por ejemplo, el servicio de alertas puede mandar un correo electrónico o un mensaje de texto al usuario que recibe una transferencia si así lo ha establecido en su perfil dentro de la entidad bancaria.

7.4 Resumen y contribuciones del capítulo

En este capítulo se han presentado las contribuciones y validaciones relacionadas con las últimas etapas del proceso de modelado de usuario que establece el marco de datos personales: generación, recuperación y aplicación de modelos de usuario. En particular, y tal y como se ha señalado en la introducción de este capítulo, el marco de datos personales permite llevar a cabo la generación de modelos de usuario e incorpora los mecanismos necesarios para su recuperación a distintos consumidores de modelos. Por otra parte, la aplicación de los modelos de usuario obtenidos del sistema es responsabilidad de los consumidores, no obstante, en el contexto de esta investigación se han realizado distintas validaciones de esta última etapa que suponen dos contribuciones originales de la tesis.

Dadas las consideraciones anteriores, las contribuciones originales que se han presentado en este capítulo son:

- Establecimiento de tres patrones principales que permiten llevar a cabo la generación de modelos de usuario considerando recursos heterogéneos y distribuidos.
- Definición de una ontología que permite representar características socio-económicas de los usuarios y obtener modelos de usuario complejos a partir de ésta.
- Diseño y desarrollo de un nuevo servicio financiero que ofrece a sus usuarios la posibilidad de recibir una cantidad determinada de dinero a través de micropréstamos de sus amigos. Para ello, se apoya en el marco de datos personales que tras un análisis de los modelos de usuario socio-económicos almacenados genera un conjunto de prestadores que satisfacen un grado de confianza, liquidez y solvencia dados.
- Identificación y validación de distintos mecanismos de generación y recuperación de conocimiento de los usuarios basados en el nuevo tipo de datos personales (eventos personales de negocio) establecido en esta investigación dentro de una arquitectura empresarial, en concreto, del ámbito bancario.

Al igual que en el capítulo anterior, las contribuciones obtenidas y las validaciones de éstas de este capítulo se enmarcan dentro de los proyectos MOET y POSDATA del Center for Open Middleware.

Cabe destacar que dentro del proyecto BUSCAMEDIA se llevó a cabo el establecimiento del patrón centralizado para la generación de modelos de usuario y su validación en el ámbito de consumo de contenidos audiovisuales. Los resultados y trabajos relacionados se incluyen en distintos entregables del proyecto, así como en los artículos (Yelmo, et al., 2011a) y (Yelmo, et al., 2010). Este último artículo fue seleccionado entre los mejores del *Second International ICST Conference, UCMedia 2010*, y su versión extendida y revisada se incluyó como capítulo de libro.

Los resultados obtenidos en este capítulo también se incluyeron dentro de distintos artículos que ya se han mencionado previamente como (San Miguel, et al., 2015d), (San Miguel, et al., 2014c) y (San Miguel, et al., 2014b).

Capítulo 8.

Gestión de identidad y privacidad centrada en el usuario

En los capítulos anteriores se describieron las etapas principales del proceso de modelado de usuario distribuido que se basa en la recuperación de recursos custodiados por proveedores de servicios externos (fuentes de datos) para producir modelos de usuario que otros proveedores (consumidores de modelos) pueden obtener para cubrir sus necesidades de negocio específicas.

Tal y como se ha venido apuntando y conforme a la cadena de valor establecida para el marco de datos personales (4.2.4 Cadena de valor), la ejecución de dicho proceso viene determinada por el usuario, ya que éste debe asumir, o al menos delegar explícitamente en el sistema, la gestión del flujo de sus recursos personales (datos y modelos de usuario) para establecer el control en el uso y la cesión de éstos. En este sentido, el presente capítulo describe la infraestructura de gestión de identidad y privacidad que permitirá llevar a cabo lo anterior.

Antes de entrar en los detalles, conviene precisar los distintos flujos de datos relacionados con el usuario e identificar aquellos que competen directamente al marco de datos personales y por lo tanto, que podrán ser habilitados para que el usuario los gestione.

La Figura 8-1 representa el diagrama de flujos de datos del marco de datos personales donde se identifican cuatro flujos que se establecen entre el usuario al que refieren éstos, las fuentes de datos externas, el propio sistema y los consumidores de modelos de usuario, así como tres procesos principales: custodia de datos, generación de modelos y aplicación de modelos.

Figura 8-1. Diagrama de flujos de los datos personales en el contexto de la tesis.

Como se observa en la figura, el usuario es fuente y sumidero de dos flujos de datos principales: “1. creación de datos” y “4. aprovechamiento modelos” respectivamente. El flujo 1 refleja que los datos utilizados en el contexto del marco de datos personales se obtienen dado que el usuario accede y utiliza los servicios provistos por las fuentes externas y por consiguiente, genera y crea datos que le refieren en éstas. En cuanto al flujo 4, éste indica que el usuario utiliza los servicios ofrecidos por los distintos consumidores de modelos y con ello, se aprovecha y beneficia de los modelos de usuario que se obtienen del marco de datos personales. La gestión de la identidad y privacidad de los flujos 1 y 4 depende principalmente de los proveedores de servicios con los que se relaciona el usuario y están fuera del alcance del marco de datos personales, sin embargo su existencia determinará la infraestructura de gestión de identidad del sistema como se verá en la sección 8.1.

Por otra parte, el flujo denominado como “2. recuperación de datos”, se establece entre las distintas fuentes de datos externas y el marco de datos personales para abastecer al sistema de los recursos necesarios en la generación de modelos de usuario. Conforme a lo expuesto en el capítulo 6, este flujo viene determinado por los servicios concretos de recuperación de datos que cada fuente habilite. Es decir, el marco de datos personales debe incorporar los mecanismos técnicos necesarios (por ejemplo, protocolos o formatos de datos) para poder establecer este flujo. Además, podrá informar al usuario sobre la existencia de estos mecanismos y los permisos que debe otorgar para su establecimiento final.

Por último, el flujo “3. recuperación de modelos” refleja cómo los modelos de usuario generados en el marco de datos personales pueden ser recuperados por distintos consumidores de modelos que los utilizan o aplican para particularizar sus servicios ofrecidos. Este flujo debe estar determinado por las preferencias e intereses de los usuarios y en este sentido, el sistema debe implantar una infraestructura para la gestión de identidad que permita establecer el flujo y por otra parte, una infraestructura para la gestión de la privacidad que permita al usuario controlarlo.

En definitiva, el marco de datos personales debe establecer **una infraestructura para la gestión de la identidad** que permita habilitar los flujos de recuperación de datos (flujo 2 en la Figura 8-1) y de recuperación de modelos de usuario (flujo 3 en la Figura 8-1) conforme a los consentimientos y las autorizaciones otorgados por el usuario. Además, lo anterior permitirá establecer una **infraestructura de gestión de privacidad** donde el usuario podrá controlar los flujos de datos habilitados y conocer los detalles de éstos.

Dadas las consideraciones anteriores, en las siguientes secciones se describen las contribuciones y validaciones de la tesis doctoral respecto a la gestión de la identidad y privacidad del marco de datos personales. En primer lugar, la sección 8.1 presenta una infraestructura para la gestión de la identidad que combina distintos modelos conceptuales de gestión y que habilitará la infraestructura de privacidad que se describe en la sección 8.2. La gestión de privacidad permitirá establecer las preferencias e intereses de los usuarios en el contexto del marco de datos personales, así como guiar a los usuarios en la toma de decisiones relacionadas con el uso y compartición de sus recursos. Seguidamente, la sección 8.3 presenta las validaciones realizadas de dichas infraestructuras en el contexto de esta investigación y por último, la sección 8.4 incluye un resumen de las contribuciones incluidas en el capítulo.

8.1 Infraestructura para la gestión de la identidad

Conforme a lo establecido en el capítulo anterior, el marco de datos personales deberá generar modelos de usuario altamente adaptativos que refieren a un usuario concreto (7.1 Características de los modelos de usuario del marco de datos personales). Esto requiere que el sistema establezca una infraestructura para la gestión de la identidad donde los datos personales y modelos de usuario describan a un usuario concreto independientemente de si estos recursos se obtienen o utilizan en dominios administrativos diferentes. Es decir, el usuario, y en concreto sus recursos personales, podría tener que ser identificados hasta en tres dominios diferentes: el de las fuentes de datos, el propio sistema y los consumidores de modelos de usuario.

En concreto, la infraestructura de gestión de identidad debe permitir:

- **Identificar a los usuarios** del sistema y **gestionar la correspondencia con sus correspondientes identidades digitales** tanto en las fuentes de datos externas y como en los consumidores de modelos de usuario. De esta forma, se podrán recuperar los datos necesarios para generar modelos de usuario altamente adaptativos y aplicarlos para el usuario concreto al que refieren.
- **Habilitar el acceso a los datos personales** de los usuarios que se encuentran custodiados por distintas fuentes de datos externas. Tal y como se vio en el capítulo 6, los mecanismos específicos que permiten la recuperación de datos dependen de la fuente considerada, sin embargo el sistema debe incluir las capacidades necesarias para obtener el consentimiento y los permisos del usuario.
- **Habilitar el acceso a los modelos de usuario generados** en el marco de datos personales a consumidores de modelos. Puesto que el sistema custodia los modelos de

usuario generados, éste puede determinar qué mecanismos habilitar para la gestión de estos recursos, optando por aquel que mejor se adapte al contexto de uso de los consumidores. Independientemente de lo anterior, los usuarios deberán dar su consentimiento y autorización para habilitar la recuperación.

La gestión de identidad puede llevarse a cabo a través de cuatro modelos conceptuales principales (3.4.1 Modelos de gestión de identidad): aislado, federado, centralizado y centrado en el usuario. El marco de datos personales se apoya en la consideración y aplicación de los tres últimos, ya que están enfocados a la gestión de identidad en contextos que promueven el intercambio y la interoperabilidad de identidades digitales entre distintos proveedores de servicios.

En los siguientes apartados se describen los modelos de gestión de identidad que pueden adoptarse en el marco de datos personales y por último, se presenta la arquitectura que permite establecer la infraestructura de gestión de identidad.

8.1.1 Modelos de gestión de identidad

Tal y como se ha señalado, el marco de datos personales establece una infraestructura para la gestión de la identidad que se apoya en tres modelos de gestión principales: centrado en el usuario, centralizado y federado.

Principalmente, el marco de datos personales establece un **modelo de gestión de identidad centrado en el usuario** en el que los usuarios pueden decidir qué identidades y recursos se vinculan en el sistema. Es decir, se considera una infraestructura que les sitúa en el centro de todos los procesos relacionados con la transferencia de sus datos personales y modelos de usuario asociados y además, aspectos relacionados con la facilidad de uso y la usabilidad.

Además del modelo anterior, el marco de datos personales se apoya en la integración de distintos **modelos de gestión centralizados** que proveedores de servicios existentes han implementado. En términos generales, estos proveedores se caracterizan por contar con un gran volumen de usuarios y ofrecer a entidades externas mecanismos de autenticación y acceso a recursos personales basados, por ejemplo, en tecnologías de autorización delegada. Dichos proveedores podrán actuar como fuente de datos externas y/o consumidor de modelos de usuario del marco de datos personales, sin embargo, dado que normalmente se trata de grandes proveedores de servicios, lo más habitual será que éstos asuman el papel de fuentes de datos externas y que el marco de datos personales adopte los mecanismos proporcionados por éstas para la recuperación de sus recursos.

Por último, para habilitar el acceso de los recursos del marco de datos personales, se han identificado dos modelos de gestión de identidad que el sistema puede adoptar: centralizado y federado. En el primer caso, el sistema actúa como un proveedor de identidad para los consumidores de modelos de usuario y así, permite que el marco de datos personales autentique a los usuarios para que autoricen a los consumidores a recuperar modelos de usuario en su nombre mediante tecnologías de delegación de autorización.

Por otra parte, se ha considerado que puede establecerse un **modelo de gestión federada** entre el marco de datos personales y nuevos proveedores de servicios que o bien pueden actuar como fuente de datos externa o bien como consumidores de modelos de usuario. En este contexto, el marco de datos personales estará sustentado por un protocolo de federación que permita a los distintos proveedores de servicios y el propio sistema intercambiar información de autenticación y autorización de forma segura.

La Figura 8-2 representa la infraestructura de gestión de identidad propuesta para el marco de datos personales donde se han identificado las entidades participantes, los modelos de gestión de identidad adoptados por el sistema para la comunicación con éstas y los distintos módulos del sistema que permiten llevarlo a cabo.

Con el objeto de facilitar la legibilidad de la figura, se ha considerado que las fuentes de datos externas adoptan un modelo de gestión de identidad centralizado. Por otra parte, se establece un modelo de gestión de identidad federado entre el sistema y distintos consumidores de modelos de usuario. Ahora bien, el marco de datos personales puede adoptar un modelo centralizado de cara a los distintos consumidores o también pueden considerarse fuentes de datos que se afilien o asocien al modelo federado de gestión de identidad (tal y como se presentará en la validación incluida en 8.3.2). Además de lo anterior, en la parte izquierda de la figura se observa el modelo de gestión centrado en el usuario que aplica el marco de datos personales.

Figura 8-2. Infraestructura de gestión de identidad en el marco de datos personales.

En definitiva, el marco de datos personales se basa en una solución general de interoperabilidad que combina distintos modelos de gestión para adaptarse a las necesidades de sus interlocutores (fuentes de datos, el usuario y consumidores de modelos de usuario). A continuación, se describen los módulos de la arquitectura que permiten llevarlo a cabo.

8.1.2 Arquitectura para la gestión de identidad

Tal y como se observa en la Figura 8-2, la infraestructura de gestión de identidad se establece en torno a cuatro módulos principales del marco de datos personales (Gestor de usuarios, Recolector, Registro y Almacén de recursos).

- **Gestor de usuarios.** Es el módulo encargado de comunicarse con el usuario y así, habilitar el modelo de gestión de identidad centrado en el usuario en el marco de datos personales. En particular, este módulo permite a los usuarios registrarse en el sistema con una nueva identidad digital, vincular las identidades de proveedores de servicios externos, autenticarse a través de distintas identidades, iniciar el proceso de autorización de recuperación de recursos o eliminar identidades previamente vinculadas. Para ello, incluye distintas interfaces de usuario usables y accesibles que guía e informa al usuario sobre los procesos asociados.
- **Registro.** Es el módulo que ejecuta las peticiones realizadas por el usuario en el módulo Gestor de usuarios, así como es el responsable de establecer las comunicaciones con los proveedores de servicios externos. Para ello, incluye dos componentes principales: un proveedor de identidad (IdP en la figura) y un servicio de descubrimiento (DS).

Las funcionalidades que lleva a cabo el módulo Registro son:

- **Alta de una nueva identidad digital.** Los usuarios pueden crear una nueva identidad digital en el sistema a través del módulo Gestor de usuario. De esta forma, el IdP del marco de datos personales recibe los datos de registro del usuario y los redirige al módulo Almacén de recursos para su almacenamiento. Este proceso requerirá la ejecución de distintos mecanismos que permitan establecer unas credenciales adecuadas para el usuario, sin embargo éstos se encuentran fuera del alcance de la tesis.
- **Vinculación identidades digitales.** El IdP recibe a través del módulo Gestor de usuarios la petición del usuario de vincular una identidad digital de un proveedor de servicios externo al sistema. Esta vinculación depende del modelo de gestión de identidad adoptado por el proveedor de servicios. Es decir, si dicho proveedor actúa como fuente de datos y proveedor de identidad central, deberán utilizarse los servicios y mecanismos provistos por éste. En el caso de que el proveedor habilite la comunicación con un protocolo de federación conforme al del sistema, la federación se realiza directamente a través de éste.

El componente IdP se apoya en el módulo Almacén de recursos para almacenar e integrar toda la información relacionada con las distintas identidades digitales del usuario y en el módulo Recolector para establecer los mecanismos y comunicaciones necesarias con aquellos proveedores que adoptan un modelo de gestión centralizado.

- **Autenticación del usuario.** El IdP lleva a cabo la autenticación del usuario conforme al mecanismo seleccionado por éste. Es decir, el usuario puede aportar sus credenciales creadas en el registro del sistema o utilizar la identidad digital de distintos proveedores de servicios vinculados previamente en el marco de datos personales.
- **Registro de recursos externos.** El módulo Registro cuenta con un componente denominado como DS que permite a distintos proveedores de servicios registrar los datos y modelos de usuario que tienen en su custodia sobre un usuario. Este registro se basa en definiciones de formatos de datos e información relacionada con la localización de los recursos, lo cual incluye los de distintas fuentes de datos externas, los del propio sistema y los del proveedor de servicios.
- **Recolector.** Bajo la solicitud del módulo Registro y en concreto, el IdP de éste, el módulo Recolector se comunica con distintas fuentes de datos externas que adopten un modelo de gestión de identidad centralizado para proceder a que el usuario autorice al sistema para acceder a sus recursos y obtener las credenciales necesarias para habilitar el acceso posterior. Esta información se envía al módulo Almacén de recursos para su uso posterior en la ejecución de los flujos de recuperación de datos.
- **Almacén de recursos.** Como se ha venido señalando, este módulo es el encargado de mantener las distintas identidades digitales de los usuarios vinculadas en el sistema, así como los permisos y las credenciales de acceso asociadas con los flujos de recuperación.

En conclusión, la infraestructura de gestión de identidad presentada se basa en un modelo de gestión centrado en el usuario que le permite registrar su identidad en el sistema, vincularla a aquella registrada en distintos proveedores de servicios para habilitar la recuperación de recursos (datos y modelos de usuarios) y llevar a cabo los procesos de autenticación y acceso a recursos externos mediante autorización delegada. Para ello, el sistema considera dos modelos de gestión de identidad que pueden adoptar los proveedores de servicios externos: centralizado y federado. Cabe señalar que el modelo federado habilitará mecanismos de *Single Sign-On* (SSO) mediante los cuales el usuario podrá utilizar la identidad del sistema o de los proveedores de servicios independientemente del dominio en el que se encuentren, mientras que en el modelo centralizado la identidad del usuario únicamente estará soportada en el dominio del proveedor de identidad.

Una vez se ha establecido la infraestructura para la gestión de identidad que permite la identificación de un usuario en distintos dominios, en la siguiente sección se presentan los detalles para gestión de la privacidad que permitirá profundizar en el papel central del usuario. Cabe señalar que la gestión de identidad y privacidad son áreas de investigación independientes, sin embargo su separación no siempre está claramente definida puesto que los recursos con los que tratan cada una de ellas son los mismos y la gestión de privacidad no puede concebirse sin el establecimiento de una gestión de identidad adecuada.

8.2 Infraestructura de gestión de la privacidad

La arquitectura para la gestión de identidad presentada en la sección anterior habilita el establecimiento de una infraestructura para la gestión de la privacidad que guía a los usuarios en la toma de decisiones sobre el uso y la compartición de sus recursos personales, así como permite establecer sus decisiones al respecto. De esta forma, se establece un modelo de gestión centrado en el usuario en el que éste se encuentra en el centro de todas las decisiones en cuanto a sus recursos personales (ver 8.1.1).

No obstante, la gestión de la privacidad implica alinear, por una parte, las preferencias establecidas por los usuarios y por otra parte, la de los proveedores de servicios que hacen uso de sus datos personales y/o modelos de usuario, incluyendo dentro de éstos las fuentes de datos externas, el marco de datos personales y los consumidores de modelos de usuario.

Dadas las consideraciones anteriores, en el contexto del marco de datos personales se han identificado tres tipos de políticas de privacidad que determinan la utilización de los recursos personales de un usuario:

1. **Preferencias de los usuarios (P0).** Este tipo de políticas reúnen las premisas establecidas por los usuarios en referencia al acceso, utilización y cesión de sus recursos personales. De esta forma, indican qué se puede o no se puede hacer con sus recursos personales. Por ejemplo, los usuarios pueden expresar que la localización o los datos derivados de ésta sólo sea utilizada por determinados consumidores de modelos de usuario o que no lo utilice nadie, incluido el marco de datos personales.
2. **Promesas del receptor de recursos (P1).** Éstas indican las condiciones y compromisos establecidos por el proveedor de servicios que va a custodiar los recursos personales del usuario. Por ejemplo, *“el número de la tarjeta de crédito no se va a compartir con ninguna entidad externa”* o *“Cuando utilizas los servicios de Google, podemos recopilar y procesar información sobre tu ubicación real”*.
3. **Políticas de gobierno de recursos (P2).** De la combinación de las políticas anteriores, se pueden establecer unas reglas para el uso y cesión de los recursos personales en cada una de las entidades que custodian éstos. Estas políticas determinan qué se puede realizar con los recursos personales de los usuarios desde el punto de vista de la entidad que los custodia e idealmente deberían reflejar las preferencias establecidas por el usuario.
4. **Políticas de solicitud de recursos (P3).** Reflejan el propósito y las condiciones de un solicitante de recursos personales respecto a una petición concreta. Un ejemplo de este tipo de políticas son: *“necesito la lista de amigos de un usuario durante 2 días”* o *“quiero saber qué canciones ha escuchado un usuario para inferir sus preferencias musicales”*.
5. **Políticas de gobierno de liberación (P4).** Éstas indican las condiciones impuestas por el proveedor de servicios que cede o libera un determinado recurso personal que tiene

en su custodia a una entidad externa. Por ejemplo, “cuando recibas el número de tarjeta de crédito se deberá borrar en 2 días”.

La Figura 8-3 representa dónde aplican las políticas de privacidad descritas anteriormente en el contexto del marco de datos personales:

Figura 8-3. Políticas de privacidad en el contexto del marco de datos personales.

En el escenario más general, se ha considerado que el usuario puede establecer a través de las herramientas provistas por distintos proveedores de servicios sus preferencias de privacidad (P0x en la figura). Dichas herramientas dependen del proveedor de servicios concreto y por ello, no siempre los usuarios podrán establecer sus preferencias en cuanto a la privacidad.

Por otra parte, cada uno de los proveedores de servicios establece una declaración de intenciones sobre el uso que va a dar a los recursos personales del usuario (políticas P1x en la figura). Como resultado de las negociaciones que pueden establecerse entre el usuario y los distintos proveedores de servicios, se trasladarán distintas políticas de gobierno en cada uno de los proveedores (P2x en la figura).

Además, en cada uno de los flujos de datos en los que se produce un intercambio de recursos personales, se pueden originar políticas de solicitud (P3x) y políticas de liberación (P4x) entre las entidades involucradas y como resultado podrán establecerse nuevas políticas de gobierno (P2x) que actualicen o determinen las establecidas previamente.

En los siguientes apartados se describen los procesos para la gestión de la privacidad desde el punto de vista del marco de datos personales y el usuario, señalando las políticas de privacidad relacionadas conforme a lo establecido.

8.2.1 Gestión de privacidad en el marco de datos personales

En el apartado anterior se han identificado las distintas políticas de privacidad que se pueden establecer entre las entidades participantes considerando el escenario más general. No obstante, es importante destacar que muchas de éstas se encuentran fuera del alcance del

marco de datos personales y por consiguiente, la infraestructura de privacidad que se presenta refiere únicamente a los recursos custodiados por el sistema y a los flujos de datos habilitados sobre los que tiene competencia. Tal y como se señaló en la introducción de este capítulo, estos flujos son principalmente los relacionados con la recuperación de modelos de usuario y la recuperación de datos, aunque éste último depende del proveedor que actué como fuente de datos externa y los servicios que incorpore.

Dadas las consideraciones anteriores, el marco de datos personales implanta un modelo de gestión de privacidad en el que las decisiones tomadas por el usuario se establecen en el sistema y se trasladan a lo largo de la cadena de consumo de información y así se lo declara a través de una política de promesa de receptor (P1b en la figura).

De este modo, el sistema incluye las herramientas necesarias para permitir al usuario establecer políticas de preferencias de privacidad (P0b en la Figura 8-3) y los mecanismos necesarios para traducir éstas a políticas de gobierno (P2b) que determinan las reglas de uso y cesión de los recursos personales que custodia. Es decir, las políticas P2b determinarán qué datos personales se pueden utilizar, qué modelos de usuario se pueden generar y cuáles de los anteriores se pueden compartir con consumidores de modelos de usuario.

Por una parte, dadas las políticas de preferencias de privacidad del usuario (P1b), el marco de datos personales determinará si se generan políticas de solicitud de acceso a los datos personales custodiados por fuentes externas (P3b). En un escenario real, estas políticas de solicitud habitualmente se realizarían a través de contratos entre las fuentes de datos externas y el sistema o incluso, pueden no establecerse y llevar a cabo la recuperación de recursos directamente a través de distintos mecanismos de recuperación como, por ejemplo, llamadas a APIs. En definitiva, el establecimiento y definición de las políticas de solicitud de datos depende de los servicios y mecanismos considerados por las fuentes de datos externas y deberán contar con el consentimiento y la autorización del usuario para poder establecerse.

Además, por otra parte, dada la solicitud de un consumidor del deseo de recuperar modelos de usuario (P3c en la figura), el marco de datos personales puede generar políticas de gobierno de liberación (P4b en la figura) que determinen las reglas de uso y cesión de los modelos de usuario solicitados y que actualizarán o generarán nuevas políticas de gobierno en el consumidor de modelos (P2c en la figura).

La Figura 8-4 refleja los módulos de la arquitectura del marco de datos personales que permiten gestionar y establecer las políticas de privacidad del usuario en el sistema. Aquí, se han detallado los distintos componentes que conforman el módulo Registro que será el encargado de llevar a cabo lo anterior en comunicación con el módulo Gestor de usuarios, el Almacén de recursos personales y el Recuperador de modelos.

Figura 8-4. Gestión e implantación de políticas en el marco de datos personales.

En la figura, las secuencias 1-4 muestran los procesos relacionados con el establecimiento de políticas de privacidad de los usuarios, la secuencia 0 indica el establecimiento de políticas de privacidad por defecto y las secuencias 5-13 el establecimiento de distintas políticas de privacidad dada la solicitud de un consumidor de modelos de usuario. A continuación se describen dichas secuencias.

8.2.1.1 Políticas de privacidad del usuario

En primer lugar, cuando un usuario se da de alta y registra en el marco de datos personales, recibe información sobre las condiciones y los compromisos del servicio ofrecido por el sistema (secuencia 1): establecer sus políticas de privacidad para la generación y compartición de modelos de usuario. De esta forma, el usuario podrá indicar sus políticas de preferencias de privacidad mediante distintas interfaces gráficas provistas por el módulo Gestor de usuarios (secuencia 2 en la figura).

En particular, el usuario cuenta con una interfaz que le permite visualizar qué recursos personales tiene vinculados en el sistema y establecer qué consumidores tienen acceso a éstos. Además, sobre dicha interfaz el usuario puede gestionar sus preferencias de privacidad respecto al uso y recuperación de datos en el marco de datos personales para determinar qué datos personales puede recuperar o qué modelos de usuario puede generar.

Principalmente, se han considerado tres tipos de permisos que el usuario puede establecer para el marco de datos personales o cada uno de los consumidores sobre sus recursos: permitido (cuando se puede hacer uso), denegado (cuando no se puede uso) e indeterminado (cuando requiere que se notifique al usuario para que decida en un determinado momento). Además, el usuario puede añadir condiciones a los permisos establecidos, por ejemplo, asociar una duración determinada al permiso concedido, solicitar una recompensa, etc.

Una vez que el usuario establece sus políticas de privacidad, éstas se reenvían al módulo Registro (secuencia 3 en la Figura 8-4) y en concreto, a un componente denominado como

Punto de administración de políticas (PAP) que será el encargado de traducir las políticas de preferencias del usuario (P0b) a políticas de gobierno de sistema P2b que se almacenarán en el módulo Almacén de recursos (secuencia 4 en la figura).

8.2.1.2 Políticas de privacidad por defecto

Además de lo anterior, se ha considerado el establecimiento de políticas de privacidad por defecto por parte del marco de datos personales (secuencia 0) que se establecen conforme a los principios de privacidad desde el diseño (3.5.1 Privacidad desde el diseño).

Estas políticas implantan niveles máximos para asegurar la protección del usuario y no permitir la compartición de sus recursos hasta que éste dé su consentimiento y autorización. En concreto, se ha asociado para los recursos de un usuario permisos “indeterminados” que pueden notificar al usuario en tiempo real o cuando éste acceda al sistema, dependiendo de las preferencias establecidas por el usuario, para obtener su consentimiento.

En definitiva, las políticas por defecto no permiten la compartición y uso de los recursos hasta no obtener el permiso del usuario. De esta forma, cuando se produce una solicitud para su respuesta se requiere que el usuario establezca su decisión.

8.2.1.3 Aplicación de políticas de privacidad

Una vez implantadas las políticas de privacidad del usuario (secuencias 1-4 en la Figura 8-4), cuando un consumidor de modelos solicita un recurso al sistema debe generar una política de privacidad de solicitud (P3c) que indique qué recurso requiere y para qué (secuencia 5). Esta solicitud se redirige al módulo Registro y en concreto, al componente llamado Punto de aplicación de políticas (PEP) (secuencia 6) que reenviará la solicitud al Punto de decisión de políticas (PDP) (secuencia 7), encargado de evaluar si la solicitud se ajusta a las preferencias del usuario para emitir un resultado. Es decir, el PDP obtiene las políticas de privacidad almacenadas del sistema (P2b) y las contrasta con la solicitud recibida (P3c) para determinar si el consumidor de modelos puede tener acceso o no a los recursos solicitados (secuencias 8-9). El resultado de esta evaluación puede tener asociada condiciones del uso de los recursos solicitados (secuencia 10). Es decir, el PDP puede permitir la recuperación del modelo de usuario solicitado pero éste puede tener asociada una política de uso en el consumidor (P2c) que se le entregaría a dicho consumidor si se permite la recuperación de los recursos solicitados (secuencia 11).

Finalmente, en función del resultado obtenido, el módulo Recuperador de modelos de usuario llevará a cabo la obtención de los recursos solicitados o no conforme a lo establecido en el capítulo 7.

Las secuencias descritas permiten implantar las políticas de privacidad del usuario. Además, cuando el resultado de evaluación de una política es indeterminado, se requiere el establecimiento de un sistema de notificación que permita obtener las preferencias de privacidad del usuario en un caso concreto.

8.2.2 Gestión de privacidad desde el punto de vista del usuario

En el apartado anterior se han descrito los detalles relacionados con el establecimiento de políticas de privacidad desde el punto de vista del marco de datos personales que considera las políticas de privacidad del usuario. Además de lo anterior, la infraestructura de privacidad debe permitir a los usuarios:

1. Conocer en cualquier momento qué recursos personales tiene asociados o vinculados en el contexto del marco de datos personales. Esto incluye los datos obtenidos de fuentes externas y los nuevos generados por el propio sistema.
2. Modificar, borrar o desvincular los valores asociados a sus recursos personales.
3. Conocer el historial de acceso y uso de sus recursos personales.
4. Establecer políticas de privacidad tal y como se ha reflejado en 8.2.1.1 Políticas de privacidad del usuario.

Para la realización de dichos requisitos se parte de la arquitectura establecida para la gestión de identidad en la sección anterior (ver Figura 8-2). A continuación, se detalla los módulos y procesos relacionados que permiten llevar a cabo los tres requisitos señalados, ya que los relacionados con el punto 4 se incluyen en el apartado anterior.

8.2.2.1 Presentación de recursos vinculados

Los usuarios a través del módulo Gestor de usuarios deberán poder obtener una visión global de los recursos personales (datos y modelos de usuario) que le refieren dentro del marco de datos personales y los valores asociados en éstos.

La Figura 8-5 muestra el diagrama de secuencia entre en las entidades participantes que permiten su realización:

Figura 8-5. Diagrama de secuencia para la presentación de recursos vinculados.

Partiendo de que el usuario se ha autenticado en el marco de datos personales a través del IdP, éste puede solicitar una visión global de todos los recursos que tienen vinculados en el sistema a través del módulo Gestor de usuarios (secuencia 1). Esta solicitud es reenviada al

módulo Registro, en concreto, a su componente DS (secuencia 2) que solicitará información sobre los recursos personales al módulo Almacén de recursos (secuencia 3).

El módulo Almacén de recursos integra la información relacionada con un usuario y en concreto, los datos personales y valores asociados de éstos custodiados en el sistema, los modelos de usuarios generados y los permisos de acceso a recursos almacenados en fuentes de datos externas. De esta forma, cuando el Registro solicita la información de un usuario, responde con los datos y modelos de usuario custodiados en el marco de datos personales y además, con los permisos que permiten la recuperación de datos custodiados por fuentes externas (secuencias 4 y 5).

Cuando el módulo Registro recibe la información del Almacén de recursos, si existen datos custodiados por fuentes externas sobre los cuales el sistema tiene acceso, los recupera. Para ello, se comunica con el módulo Recolector (secuencia 6) que obtiene los datos y valores solicitados de distintas fuentes externas para devolvérselo al módulo Registro (secuencias 7-10).

Finalmente, el módulo Registro devuelve todos los datos, modelos de usuario y valores asociados al Gestor de usuarios que se los presenta al usuario (secuencias 11-12).

8.2.2.2 Modificación de los recursos

Dada la presentación de los recursos anteriores (datos y modelos de usuario, así como los valores asociados a éstos), el usuario puede modificar, borrar o desvincular los valores asociados.

En concreto, el usuario podrá borrar o modificar recursos personales custodiados por el marco de datos personales y por otra parte, desvincular o denegar el acceso a recursos custodiados por fuentes de datos externas al sistema. Para esto último, se eliminan los permisos de acceso asociados en el sistema.

Para la realización de lo anterior, se establece el siguiente diagrama de secuencia:

Figura 8-6. Diagrama de secuencia para la modificación/borrado recurso personal.

El indica que quiere modificar un determinado recurso usuario a través del módulo Gestor de usuario y éste lo redirige al módulo Registro (secuencias 2-3). El módulo Registro se comunica con el Almacén de recursos para que éste modifique el recurso (secuencia 3-4). El módulo Almacén responde con el resultado de la operación y ésta se le comunica al usuario (secuencias 5 a 7).

En el diagrama se ha representado la modificación de un recurso pero ésta es la misma cuando un usuario quiere borrar o denegar el acceso a un determinado recurso, ya que lo único que cambia son las operaciones asociadas.

8.2.2.3 Historial de uso de un recurso

Además de lo anterior, un usuario puede solicitar el historial de uso de los recursos asociados en el sistema. Dado que el Almacén de recursos inscribe cada una de las peticiones realizadas por el resto de módulos y entidades, la obtención del historial de uso es igual a la presentada anteriormente, pero la operación solicitada al Almacén recursos es la recuperación de los datos de acceso:

Figura 8-7. Diagrama de secuencia para la consulta de historial de uso.

En relación a este proceso, también puede considerarse que cada una de las entidades que almacenen recursos de los usuarios lleven un registro del uso de sus recursos y por lo tanto, para la presentación del historial será necesario interactuar con dichas entidades y recuperar los datos del historial tal y como se presentó en la Figura 8-5 Diagrama de secuencia para la presentación de recursos vinculados.

8.3 Validaciones de las infraestructuras de gestión de identidad y privacidad

En esta sección se describen las validaciones realizadas de las infraestructuras de gestión de identidad y privacidad presentadas en las secciones anteriores. En primer lugar, se presenta una validación relacionada con la gestión de distintas identidades digitales de un usuario en servicios de redes sociales. Por otra parte, en el apartado 8.3.2 se incluye la descripción de un controlador de privacidad desarrollado para el establecimiento de políticas de privacidad de los usuarios en un dominio federado.

8.3.1 Gestión de identidades digitales en servicios de redes sociales

Dado el número creciente de proveedores de servicios que adoptan soluciones de gestión de identidad centralizada para ofrecer a entidades externas funcionalidades relacionadas con la autenticación y el acceso a recursos personales de sus usuarios, la integración de estos modelos ha sido abordada desde distintas perspectivas. En este sentido, existen soluciones

comerciales, como las ofrecidas por Janrain o Gigya, que permiten integrar más de veinte modelos de gestión de identidad diferentes para llevar a cabo la autenticación a través de éstos. Además, han aparecido distintas iniciativas fruto de acuerdos entre distintas instituciones y empresas como, por ejemplo OpenSocial y el proyecto DataPortability⁸⁸, que tratan de unificar los mecanismos de recuperación y acceso a los recursos de identidad de un usuario para la facilitar la gestión de identidad en distintos dominios.

En el contexto de esta investigación no se ha profundizado sobre las soluciones de integración existentes. No obstante, se ha llevado a cabo una validación de los elementos necesarios de la infraestructura de gestión de identidad del marco de datos personales descrita en la sección 8.1 para integrar distintos modelos de gestión centralizados.

La validación mencionada se ha llevado a cabo en el contexto del servicio de pago Posdata presentado en 6.4.1 Recuperación de datos personales en servicios de mensajería y redes sociales: el servicio de pago Posdata. Este servicio permite a sus usuarios realizar pagos a través de distintos canales de comunicación, como servicios de mensajería (por ejemplo, Telegram) o redes sociales (como Facebook o Twitter) mediante mensajes propios de cada canal y la realización final de los pagos a través de los saldos asociados a la tarjeta TUI.

Lo anterior requiere establecer una infraestructura para la gestión de identidad que permita a los usuarios vincular en el servicio Posdata las identidades digitales de los canales de comunicación que desee habilitar y del proveedor de la tarjeta TUI. En concreto, dicha infraestructura se compone de los siguientes elementos:

1. **Interfaz de usuario:** informa sobre las identidades digitales que el usuario puede asociar o vincular al servicio. De esta forma, cuando un usuario desea vincular una identidad en el servicio se le informa de los pasos que debe llevar a cabo. En particular, se han utilizado las soluciones ofrecidas por los servicios de redes sociales Facebook y Twitter pero puede extrapolarse a otros proveedores de servicios que cuentan con soluciones similares.
2. **Registro:** dada la solicitud de un usuario para vincular una identidad digital en el servicio, establece la comunicación con el proveedor seleccionado e implementa los mecanismos provistos por éste. En este sentido, se han utilizado las APIs de Facebook y Twitter para obtener las claves de acceso a los recursos personales de los usuarios. Dichas APIs se basan en el protocolo OAuth conforme a lo presentado en el capítulo 6.
3. **Almacén de recursos:** custodia la información de identidad digital del usuario en distintos proveedores de servicios. De esta forma incluye los identificadores de los usuarios en cada uno de los dominios considerados (el servicio de pago Posdata, los servicios que actúan como canales de comunicación y el proveedor de la tarjeta TUI) y las claves de autorización asociadas a dichas identidades que permiten la recuperación de recursos custodiados por éstos.

En definitiva, el servicio de pago Posdata se sustenta en una infraestructura de gestión de identidad basada en la integración de distintos modelos de gestión centralizada que permitirá

⁸⁸ DataPortability Project, <http://dataportability.org/>

la autenticación en el servicio a través de las identidades digitales y la recuperación de recursos generados en dicho servicio. Los detalles de implementación concretos de dicha infraestructura se han descrito en capítulo 6.

8.3.2 Controlador de privacidad para un dominio federado de gestión de identidad

Además de lo anterior, en el contexto de esta investigación se ha validado la infraestructura de gestión de privacidad presentada en la sección 8.2, considerando el establecimiento de un modelo federado de gestión de identidad entre las distintas fuentes de datos externas, el marco de datos personales y los consumidores de modelos de usuario. Las validaciones realizadas se han llevado a cabo conforme a las especificaciones del proyecto Liberty Alliance, aunque puede ser extrapolado a otras soluciones de federación.

Tal y como indica la Figura 8-8, por una parte, los distintos proveedores de servicios participantes (las fuentes de datos externas, el marco de datos personales y los consumidores de modelos de usuario) establecen un dominio de confianza basado en una gestión de identidad federada y por otra parte, sobre éste dominio se permite la gestión de la privacidad al usuario a través de un componente del marco de datos personales denominado como **Controlador de privacidad** (*Privacy Controller – PC*).

Figura 8-8. Controlador de privacidad en un dominio federado.

En el modelo federado se establece un servicio de descubrimiento (del inglés, *Discovery Service – DS*) que permite a los distintos proveedores de servicios registrar los recursos de identidad que custodian y que está gestionado por el marco de datos personales. En concreto, el DS actúa como una entidad central que integra punteros a los datos personales y modelos de usuario custodiados por las fuentes de datos externas, el marco de datos personales (cuyos

recursos se encuentran en el módulo Almacén de recursos) y los consumidores de modelos usuario.

Dadas las consideraciones anteriores, se ha diseñado y desarrollado un PC que permite a los usuarios:

1. Obtener una presentación y visión global de los recursos custodiados por los distintos proveedores de servicios del dominio federado. En concreto, la información que se muestra al usuario es: los datos personales almacenados en fuentes externas, los modelos de usuario generados por el marco de datos personales (almacenados en el módulo Almacén de recursos) y los modelos de usuario que distintos consumidores pueden haber obtenido del sistema.
2. Modificar los valores asociados a los recursos asociados en el dominio federado, así como la eliminación de éstos.
3. Consultar el historial de uso de los recursos intercambiados entre los distintos proveedores del dominio federado. Este historial permitirá saber al usuario quién ha accedido a qué recursos y qué operaciones se han realizado sobre éstos.
4. Establecer políticas de privacidad que gobiernen la información custodiada por cada uno de los proveedores de servicios del dominio federado.

Para realizar lo anterior, se requiere que el PC esté autenticado en nombre del usuario frente al dominio federado para poder interactuar con el resto de entidades. Dicha autenticación se puede llevar a cabo conforme a distintos mecanismos descritos en el Liberty ID-FF o Liberty ID-WSF y como resultado, el PC obtendrá un puntero de referencia (EPR) al DS y los credenciales para el acceso a éste.

A continuación, se describe el detalle de los procesos que permiten llevar a cabo la gestión de identidad al usuario, identificando los componentes participantes y las relaciones que se establecen entre estos.

8.3.2.1 *Presentación de recursos*

La Figura 8-9 representa el diagrama de secuencia que se establece en el marco de datos personales para poder ofrecer al usuario una visión global de los recursos personales que tiene asociados dentro del dominio federado.

Figura 8-9. Diagrama de secuencia para la presentación de recursos en dominio federado.

En primer lugar, el usuario solicita en el PC una visión global de sus recursos asociados (secuencia 1). Seguidamente, el PC solicita al DS todos los recursos de identidad del usuario dentro del dominio federado (secuencia 2) y éste tras obtener la información solicitada (secuencia 3) responde con la localización de cada recurso y los credenciales de acceso a éstos (secuencia 4). Con esta información, el PC presenta al usuario una visión global de sus recursos agrupados conforme a los distintos proveedores de servicios que los custodian (secuencia 5). La implementación para el prototipo desarrollado de las secuencias 1-3 se basan en la especificación *Liberty ID-WSF Discovery Service Specification*.

8.3.2.2 Modificación y borrado de los valores asociados

Dada la presentación global obtenida, el usuario puede solicitar al PC los valores asociados a los recursos custodiados por un determinado proveedor de servicios y modificar o eliminar éstos. De esta forma, se establece el diagrama representado en la siguiente figura:

Figura 8-10. Diagrama de secuencia para la visualización de valores asociados a un recurso y su modificación.

Primero, el usuario selecciona el recurso sobre el que quiere obtener su valor asociado (secuencia 1). El PC solicita al proveedor de servicios que almacena el recurso seleccionado su valor asociado aportando el EPR obtenido en el proceso de autenticación y los credenciales de acceso obtenidos en el proceso anterior (secuencia 2). Este proveedor podrá ser cualquiera de los identificados en la Figura 8-8: fuentes de datos externas, el módulo Almacén de recursos del marco de datos personales y los consumidores de modelos de usuario.

Cuando un proveedor de servicios recibe la solicitud del PC, comprueba los credenciales de acceso y las políticas de privacidad establecidas sobre el recurso solicitado (secuencia 3) y conforme a esto responde al PC (secuencia 4). Si el PC tiene acceso al recurso solicitado, se entrega el valor solicitado y éste se lo presenta al usuario (secuencia 5). La comunicación entre el PC y el proveedor de servicios se lleva a cabo conforme al protocolo *Liberty Data Service Template* (DST).

Una vez que el usuario obtiene el detalle sobre los recursos solicitados, éste puede editar o borrar los valores asociados a través de la interfaz de usuario del PC (en la parte izquierda de la Figura 8-10 se presenta un ejemplo de dicha interfaz). Dichas operaciones están soportadas por el protocolo DST de Liberty y su realización se lleva a cabo directamente entre el PC y el proveedor de servicios (secuencias 6 y 7).

8.3.2.3 Presentación del historial de uso

Además de las funcionalidades anteriores, el usuario puede solicitar al PC el historial de uso de un determinado recurso y así, conocer quién ha accedido a qué recursos y en qué momento. Además, se puede incluir otra información como, por ejemplo, el motivo del uso, las políticas asociadas a éste o el resultado de las solicitudes de acceso o uso a dicho recurso, etc.

La siguiente figura muestra un ejemplo del historial de uso del número de tarjeta de crédito de un usuario, indicando qué entidades (WSC en la figura) han solicitado dicho recurso y para qué (Action en la figura). Además, se indica la fecha de solicitud (Date) y el resultado de la petición (Access).

segur@

Privacy Controller

myPaymentBroker's history

ATTRIBUTE	VALUE	HISTORY			
CreditCard	<div></div>	WSC	ACTION	DATE	ACCESS
		goShopping	request	Wed Jul 08 13:46:36 CEST 2009
		myPaymentBroker	request	Wed Jul 08 13:46:42 CEST 2009
		goShopping	request	Wed Jul 08 15:46:36 CEST 2009

ERICSSON

Back

Figura 8-11. Interfaz del controlador de privacidad con el historial de uso de la tarjeta de crédito de un usuario.

Para la realización de lo anterior, el PC puede apoyarse en distintas especificaciones como, por ejemplo, el *Liberty ID-WSF Accounting Service* o el mecanismo de *Subscription/Notification* del proyecto Liberty. Además, en el contexto de esta investigación, se ha propuesto extender el protocolo DST de Liberty para que cada proveedor de servicios reporte sobre las solicitudes y usos realizados a los recursos que custodian.

8.3.2.4 Establecimiento de políticas de privacidad

Los usuarios pueden establecer políticas de privacidad que reflejen sus preferencias en cuanto al uso y la compartición de sus recursos almacenados en distintos proveedores de servicios a través del PC. Para ello, incluye una interfaz de usuario (como por ejemplo, la indicada en Figura 8-12) donde se presenta para cada recurso los proveedores de servicios que pueden requerir su acceso y distintos niveles de permisos que pueden asociarse para cada proveedor: “permitir” “denegar” o “preguntarme” (lo cual requiere notificar al usuario).

The screenshot shows a web interface titled "Privacy Controller" for a user named "segur@". The main heading is "myPaymentBroker's policy". Below this is a table with three columns: "ATTRIBUTE", "REGISTERED SERVICES", and "PERMISSION". The "ATTRIBUTE" column has a single entry "CreditCard". The "REGISTERED SERVICES" column lists "goShopping", "myPaymentBroker", and "Others". The "PERMISSION" column for each service has three radio button options: "Allow", "Deny", and "Ask Me". For "goShopping", "Allow" is selected. For "myPaymentBroker", "Allow" is selected. For "Others", "Deny" is selected. At the bottom of the interface, there are logos for "UPM" and "ERICSSON", and two buttons: "Change Policy" and "Back".

ATTRIBUTE	REGISTERED SERVICES	PERMISSION
CreditCard	goShopping	☒ Allow ☐ Deny ☐ Ask Me
	myPaymentBroker	☒ Allow ☐ Deny ☐ Ask Me
	Others	☐ Allow ☒ Deny ☐ Ask Me

Figura 8-12. Ejemplo interfaz de usuario para el establecimiento de políticas de privacidad.

Una vez que el usuario establece su política de privacidad, el PC traduce ésta a una política de gobernanza que entrega al proveedor que custodia dicho recurso y que será el encargado de gestionar y evaluar las solicitudes a los recursos custodiados conforme a dicha política.

El PC considera la definición de políticas de privacidad por defecto en las que se deniega el acceso a los recursos asociados con un usuario conforme a lo establecido en los principio de Privacidad desde el diseño.

La evaluación de las políticas de privacidad se lleva a cabo conforme a la arquitectura presentada para XACML en el estado del arte.

8.4 Resumen y contribuciones del capítulo

En este capítulo se han presentado las infraestructuras de gestión de identidad y privacidad que sustentan el marco de datos personales para llevar a cabo el proceso de modelado de usuario distribuido conforme a las preferencias e intereses de privacidad del usuario. El establecimiento de dichas infraestructuras ha requerido determinar los flujos de los recursos personales de los usuarios e identificar aquellos sobre los cuales el sistema tiene competencia para habilitar su gestión al usuario.

En particular, las contribuciones incluidas en este capítulo son:

- Análisis de los flujos de datos que se establecen en torno al marco de datos personales e identificación de aquellos sobre los cuales el sistema tiene capacidad para determinar su ejecución y habilitar su gestión al usuario.
- Propuesta de una infraestructura para la gestión de identidad centrada en el usuario que permite a éste la administración de modelos de gestión centralizados y federados establecidos entre distintos proveedores de servicios.
- Propuesta de una infraestructura para la gestión de la privacidad que se basa en el establecimiento de políticas de privacidad conforme a los intereses y preferencias del usuario y distintas funcionalidades que guían al usuario sobre la toma de decisiones en el uso y compartición de sus recursos personales.

Las contribuciones anteriores han sido validadas en el contexto de dos proyectos de investigación fruto de la colaboración con las empresas Ericsson y el Banco Santander respectivamente y esto ha dado lugar una serie de méritos de investigación, entre los que se destacan:

- Una patente en explotación por parte de la empresa Ericsson con título *“Method for Privacy Management in an Identity Network, Physical Entities and Computer Programs Therefor”* (Monjas, et al, 2010). Ésta se basa en la infraestructura de privacidad propuesta en este capítulo y en particular, su validación para un dominio federado de gestión de identidad.
- Con respecto a la patente mencionada, se han obtenido distintos reconocimientos como el Premio Patentes e Inventores de la empresa Ericsson y ésta ha sido extendida para su protección en países como, por ejemplo, Australia, Japón o Estados Unidos.
- La publicación del artículo titulado *Self-Service Privacy: User-Centric Privacy for Network-Centric Identity* (del Álamo, et al. 2010) que presentó la autora de esta tesis en el congreso internacional *International Conference on Trust Management (IFIPTM 2010)*.
- Publicación de los artículos (San Miguel, et al., 2014a), (San Miguel, et al., 2014c), (San Miguel, et al., 2015a) y (San Miguel, et al., 2015a) citados en capítulos anteriores y que incluyen los detalles relacionados con la infraestructura de gestión de identidad centrada en el usuario y la integración de modelos de gestión centralizados provistos por servicios de redes sociales.

Capítulo 9.

Validación y resultados

En este capítulo se recogen las distintas acciones que se han llevado a cabo para validar el enfoque global y las contribuciones originales que se han ido señalando a lo largo del documento. Cada una de estas acciones se ha ejecutado dentro de diferentes proyectos de investigación con empresas y actores de relevancia donde se ha participado activamente en los resultados de los mismos. En particular, se han realizado distintos prototipos y tareas propias de colaboración con el resto de socios del proyecto como la elaboración de documentos, asistencia a reuniones de coordinación, audioconferencias, presentación en sesiones plenarias y auditorías de evaluación, etc.

Además, lo anterior se ha complementado con trabajos de soporte académico para la realización de dos proyectos fin de carrera y otras actividades de formación de alumnos de la ETSIT que han sido tutelados por la autora de la tesis doctoral.

Fruto de los trabajos realizados se han obtenido distintos méritos de investigación como la protección intelectual e industrial mediante dos patentes, la publicación de un artículo en una revista indexada en el JCR y distintos artículos que han sido presentados en congresos internacionales y nacionales de relevancia. Cabe señalar que algunos de estos resultados han sido galardonados con premios de distintas instituciones, así como en las conferencias donde han sido presentados.

A continuación, en la sección 9.1 se describen los proyectos de investigación que han sido receptores de los resultados de la tesis y que han permitido la validación de éstos, así como se describen otros trabajos académicos en los que ha participado la autora. Posteriormente, en la sección 9.2 se incluyen los resultados de investigación obtenidos, indicando las patentes, publicaciones y premios otorgados. Finalmente, la sección 9.3 incluye un resumen del capítulo.

9.1 Validaciones en proyectos de investigación y otros trabajos académicos

Como se ha señalado anteriormente, las contribuciones originales de la tesis han sido validadas dentro cuatro proyectos de investigación y mediante trabajos de soporte académico. A continuación, se detallan éstos.

9.1.1 Validación en el proyecto Segur@

El proyecto Segur@ es un proyecto de investigación nacional enmarcado en la tercera convocatoria del Programa CENIT de la iniciativa Ingenio 2010, cofinanciado por el Centro para el Desarrollo Tecnológico Industrial (CDTI). El proyecto, liderado por Telefónica I+D, ha estado constituido por doce empresas punteras de seguridad TIC como Ericsson, Ingeniería de Sistemas para la Defensa de España (ISDEFE), J&A Garrigues, Atos Origin, Safelayer Secure Communications, Secuware, etc., además de quince organismos públicos de investigación de relevancia en el sector entre los que destacan el Consejo Superior de Investigaciones Científicas (CSIC), la Universidad Politécnica de Madrid, la Universidad Carlos III de Madrid o la Fundación ROBOTIKER, entre otros.

El grupo de Sistemas de Tiempo Real y Arquitectura de Servicios Telemáticos de la Universidad Politécnica de Madrid, al que pertenece la autora de esta tesis, participó en el proyecto trabajando de forma directa con la empresa Ericsson España dado un convenio de colaboración entre ambos.

El objetivo general del proyecto era generar un marco de confianza y seguridad para el uso de las TIC en la sociedad de la información, lo cual tiene vinculación directa con las contribuciones de esta tesis doctoral presentadas y en concreto, con las incluidas en el Capítulo 4 Nuevos modelos de negocio para el ecosistema de datos personales y el Capítulo 8 Gestión de identidad y privacidad centrada en el usuario.

Principalmente, los trabajos relacionados con la tesis doctoral se enmarcaron dentro del paquete de trabajo 5 (Entorno tecnológico para la gestión integral de la identidad digital) en la tarea 5010 (Arquitectura de gestión de la identidad y la privacidad) y el paquete de trabajo 6 (Entorno tecnológico para la gestión de la seguridad) en la tarea 6010 (Análisis de componentes y coyuntura). Dentro de éstas se realizaron los siguientes trabajos relacionados con esta tesis doctoral:

- Definición de un nuevo modelo de negocio basado en la intermediación de información de los usuarios para la creación de anuncios en redes sociales. Lo anterior incluye el diseño de una arquitectura para distribuir información en distintas redes sociales con un punto de control centralizado en el usuario. Los resultados obtenidos se incluyeron en el entregable *“E5004 Entorno Tecnológico para la Gestión Integral de la Identidad Digital”* de la última anualidad del proyecto y como se verá en 9.2 Resultados obtenidos, originó una patente. En el Capítulo 4 de este documento, y en concreto, en 4.3.1 Modelo de negocio de intermediación de publicidad en redes

sociales, se han incluido los resultados más relevantes obtenidos de estos trabajos que han supuesto la validación y particularización del modelo de negocio propuesto en esta tesis doctoral, así como de la arquitectura de referencia descrita en el Capítulo 5.

- Búsqueda y selección de nuevas fuentes de información de seguridad vinculadas a la interacción de los usuarios y sus dispositivos electrónicos. Además, se llevó a cabo la definición de un modelo de datos para unificar la representación de dichas fuentes de información. Los resultados obtenidos se englobaron en el entregable “E6001 Fuentes de Información en Herramientas de Gestión de Seguridad”. Lo anterior permitió la identificación y caracterización de distintas fuentes de datos que se han reflejado en el Capítulo 6 de este documento.
- Diseño de una arquitectura para establecer el control centralizado por parte de los usuarios sobre sus datos personales que pueden estar dispersos en distintos nodos en una infraestructura de gestión de identidad federada. El trabajo se complementó con un análisis de las características necesarias para mejorar la experiencia de usuario en el campo de la privacidad de los datos personales, y con un estudio de cómo la información semántica podría mejorar la funcionalidad de la herramienta. Los resultados obtenidos se incluyeron dentro del entregable “E5004 Entorno Tecnológico para la Gestión Integral de la Identidad Digital” y originó una patente. En el Capítulo 8 de este documento, y en particular, en 8.3.2 Controlador de privacidad para un dominio federado de gestión de identidad, se han sintetizado los resultados más relevantes de estos trabajos que suponen una validación de la infraestructura de privacidad propuesta en esta tesis doctoral.
- Diseño e implementación de un módulo de generación de políticas de privacidad por parte del usuario que permite verificar qué servicio está disponible en cada red social, cómo es usada y cómo se puede modificar, actualizar o borrar. Los resultados obtenidos también se incluyeron en el entregable “E5004 Entorno Tecnológico para la Gestión Integral de la Identidad Digital”. Los trabajos realizados permitieron validar la infraestructura de gestión de identidad y privacidad propuesta en esta tesis doctoral (Capítulo 8) para el ámbito de redes sociales.

Finalmente, cabe señalar que Ericsson España tenía como objetivo específico la generación de resultados de investigación que pudieran incorporar directamente a sus productos y servicios a través de sus unidades de negocio. En este sentido, se elaboraron distintas propuestas de ideas innovadoras patentables (*Invention Disclosures*) que tras ser analizadas y revisadas originaron su protección mediante dos patentes tal y como se describirá en 9.2.1 Protección intelectual e industrial a través de patentes.

9.1.2 Validación en el proyecto Buscamedia

El proyecto Buscamedia (Buscamedia, 2015) fue un proyecto CENIT cofinanciado por el CDTI que se ejecutó en el periodo 2009-2012 con un consorcio de doce empresas y ocho organismos públicos de investigación que cubren cada uno de los puntos de la cadena de valor del proyecto, aportando un equipo multidisciplinar de varias ramas de la ciencia tecnológica y

la lingüística aplicada. Entre los participantes destacan grandes empresas como Atos Origin, Activa Multimedia o INDRA, PYMES especializadas en distintas áreas del sector audiovisual y en soluciones ontológicas para contenidos multimedia y multilingües como DAEDALUS, Fractalia o BMAT y distintos organismos de investigación de relevancia como la Universidad Politécnica de Madrid, la Universidad Carlos III de Madrid, Tecnalia o la Fundación CTIC. Los trabajos de este proyecto relacionados con la tesis doctoral son fruto de la colaboración directa con la empresa GFI Informática.

El objetivo principal del proyecto Buscamedia era la consecución de un buscador multimedia semántico que pudiera adaptarse dinámicamente a cualquier red, terminal, contexto y usuario. Para ello se colaboró en distintas tareas relacionadas directamente con la tesis doctoral y se realizaron prototipos y pruebas de concepto que permiten validar el enfoque global de la misma. En particular, los trabajos realizados en el proyecto se centraron principalmente en la recuperación de recursos, su anotación y almacenamiento para el establecimiento de un patrón de generación de modelos de usuario centralizado (Capítulo 7). Además, se trabajó en aspectos relacionados con la gestión de recursos dentro del buscador desarrollado en el proyecto, lo cual tiene relación con los resultados descritos en el Capítulo 8.

En concreto, los trabajos realizados dentro del proyecto fueron:

- Análisis e identificación de las ontologías existentes para la representación de datos personales de los usuarios. Los resultados obtenidos se incluyeron en el entregable *“D7.1.1 Personalización y adaptación al contexto de usuario”* y han sido sintetizados en el Capítulo 3 Estado del arte de este documento.
- Diseño e implementación de un programa araña especializado en la recuperación de datos personales públicos de los usuarios en redes sociales y en su anotación para facilitar la generación de modelos de usuario. Los resultados obtenidos se llevaron a cabo en el paquete de trabajo 4 (Anotación e indexación semántica de archivos multimedia) y se incluyeron en los entregables *“D4.1.x Algoritmo de anotación e indexado de texto”* del proyecto. Estos trabajos suponen una validación de la recuperación de datos personales públicos en el marco de datos personales descrita en el Capítulo 6 de este documento.
- Diseño de un espacio de trabajo dirigido a usuarios finales que permite gestionar los datos personales y contenidos multimedia relacionados con el buscador de Buscamedia. Este espacio de trabajo permite monitorizar los datos de los usuarios disponibles así como la experiencia de uso del buscador. Los trabajos relacionados con esta línea de investigación se enmarcaron en los paquetes de trabajo 7 (Personalización y adaptación al contexto de usuario) y 8 (Nuevas tecnologías de acceso) del proyecto y se incluyeron en los entregables *“D7.1.x Personalización y adaptación al contexto de usuario”* y *“D8.3.x Presentación adaptada al usuario e interfaces en distintas anualidades”*. Además, los resultados obtenidos han contribuido a definición de las infraestructuras de gestión de identidad y gestión de privacidad descritas en el Capítulo 8 de este documento.

- Implementación de una prueba de concepto que se integró en el espacio de trabajo mencionado anteriormente y que se basa en la personalización de contenidos de redes sociales en función de las características particulares de los usuarios y los contenidos que éstos consumen y visualizan a través de Buscamedia. Al igual que en el caso anterior los resultados obtenidos forman parte de los entregables D7.1.x y D8.3.x. Estos trabajos se llevaron a cabo mediante técnicas de modelado de usuario basadas en tecnologías semánticas y permitieron validar el enfoque centralizado de generación de modelos de usuario descrito en el Capítulo 7 de este capítulo.

Los resultados obtenidos en el proyecto Buscamedia han permitido establecer distintas contribuciones de la tesis doctoral, así como adquirir conocimiento sobre distintas fuentes de datos, el proceso de modelado de usuario y tecnologías semánticas. Todo lo anterior se ha podido extrapolar a otros ámbitos de aplicación como el sector financiero que ha sido analizado dentro del Center for Open Middleware tal y como se describirá a continuación.

9.1.3 Validación en el Center for Open Middleware

El Center for Open Middleware (COM) es un centro de tecnológico mixto creado por la Universidad Politécnica de Madrid y el Banco Santander en colaboración con sus divisiones tecnológicas ISBAN –Ingeniería de Software Bancario y PRODUBAN –Servicios Informáticos Generales. El centro tiene distintas áreas de investigación y proyectos relacionados con el sector financiero y necesidades específicas de la infraestructura del Banco Santander. La autora de esta tesis doctoral ha participado en dos proyectos de investigación a los que ha aportado resultados de esta investigación. A continuación se describen dichos proyectos.

9.1.3.1 Proyecto MOET

El proyecto MOET (Middleware Orientado a Eventos) se ejecutó durante los años 2012 y 2013 dentro de la línea de investigación de Arquitecturas Abiertas de Middleware del COM con el objeto de investigar, analizar y evaluar las posibilidades que el paradigma de arquitecturas EDA ofrece en distintos entornos del Banco Santander. Para ello el proyecto se llevó a cabo con la colaboración de personal de ISBAN.

Específicamente, el proyecto investigó la generación y uso de eventos de negocio en tiempo real en el contexto de aplicaciones basadas en la arquitectura BankSphere del Banco Santander. Conjuntamente, se desarrolló un entorno de experimentación basado en una arquitectura EDA con distintas herramientas de código libre que facilitó la implementación de una prueba de concepto basada en la generación de eventos relacionados con los usuarios para su explotación por parte de terceras entidades. Asimismo, el entorno de desarrollo permitió la realización de distintas pruebas de rendimiento.

Los resultados obtenidos en el proyecto se incluyeron en los entregables internos: *“Estado del Arte”*, *“Propuesta de Arquitectura”* y *“Testbed para la experimentación de arquitecturas EDA y su integración en BKS. Validación y análisis de prestaciones”*. Además, los desarrollos software se empaquetaron en distintas máquinas virtuales para la evaluación de los socios del proyecto.

Los trabajos del proyecto MOET permitieron validar las contribuciones relacionadas con el paradigma de arquitecturas EDA, analizando su viabilidad, eficiencia y las posibilidades de ésta en el marco general de la tesis. En concreto, dichos trabajos suponen una validación de la recuperación de un nuevo tipo de datos personales identificado en esta investigación tal y como se ha reflejado en el Capítulo 6. Además, han permitido establecer un patrón para la generación de modelos de usuario desacoplado que se ha descrito en el Capítulo 7, junto con su validación en el ámbito bancario.

9.1.3.2 Proyecto POSDATA

El proyecto POSDATA se ejecutó en los años 2014 y 2015 y estaba incluido dentro de la línea de investigación de Ecosistemas para el Intercambio de Datos de Carácter Personal y Social del COM. El objetivo de proyecto era proporcionar un marco conceptual y el soporte tecnológico necesario para permitir el análisis experimental de la información personal y social de los titulares de la tarjeta universitaria TUI del Banco Santander con el propósito de monitorizar y mejorar el uso de los servicios, personalizar la oferta de servicios vinculados a la TUI y ampliar el catálogo de servicios para la comunidad universitaria. En este sentido, principalmente se llevaron a cabo las siguientes actividades relacionadas con la validación del enfoque global de la tesis doctoral:

- Diseño y desarrollo de nuevos servicios financieros. En particular, se implementó una aplicación de pago directo entre personas a través de la red social Facebook que permite validar la recuperación de datos personales públicos y privados conforme a lo descrito en el Capítulo 6 (6.4.1 Recuperación de datos personales en servicios de mensajería y redes sociales: el servicio de pago Posdata). Además, se desarrolló un servicio de micropréstamos entre amigos que moviliza la red social de un usuario para recibir una determinada cantidad de dinero y que permite validar la recuperación y aplicación de modelos de usuario como se detalló en el Capítulo 7 (7.3.1 Patrón centralizado para generación de modelos de usuario socio-económicos).
- Definición de un modelo semántico de datos elaborado a partir de los perfiles de usuario en Facebook y de las transacciones económicas realizadas entre usuarios. En el Capítulo 7 (7.3.1.1 Modelo de usuario interoperable) se han incluido los detalles más relevantes de dicho modelo.
- Diseño e implementación de una plataforma que permite el almacenamiento, descubrimiento, acceso y uso de modelos de usuario a medida a distintos proveedores de servicios a través de tecnologías semánticas. Este trabajo supone una validación del patrón centralizado para la generación de modelos de usuario que se ha descrito en el Capítulo 7 (7.3.1 Patrón centralizado para generación de modelos de usuario socio-económicos).
- Análisis de nuevos modelos de negocio en el entorno bancario y aquellos relacionados con el intercambio de datos. Los resultados obtenidos se han sintetizado en el Capítulo 3 Estado del arte (3.6 Modelos de negocio en el ecosistema de datos personales

actual) y en el Capítulo 4 Nuevos modelos de negocio para el ecosistema de datos personales.

- Propuesta de un nuevo modelo de negocio para la promoción de nuevos servicios financieros basado en un ecosistema de intercambio de datos personales centrado en el usuario. Los resultados obtenidos se han descrito en el Capítulo 4 (4.3.2 Modelo de negocio para la promoción de nuevos servicios financieros).

Además de los trabajos técnicos mencionados, dentro del proyecto se llevó a cabo la extrapolación y abstracción de los conocimientos adquiridos en el resto de proyectos para proponer un marco de datos personales de referencia. En concreto, en el Capítulo 4 se presenta el modelo de negocio propuesto en esta investigación y el Capítulo 5 la arquitectura de referencia que se sustenta en éste, ambos realizados en el contexto de este proyecto.

Los resultados obtenidos se incluyeron en los entregables internos del proyecto: *“Visión del proyecto: hacia un nuevo ecosistema de datos personales”*, *“Informe de estado del arte”*, *“Modelo de usuario interoperable y nuevas aplicaciones”* y *“Arquitectura de un marco de datos personales”*.

9.1.4 Validación con soporte de trabajos académicos

Durante la realización de esta tesis doctoral se colaboró en distintos trabajos del contexto académico cuyos resultados han contribuido al enfoque global de validación con distintas pruebas de concepto y estudios de viabilidad de tecnologías y prestaciones. Estas colaboraciones son: dos proyectos fin de carrera y distintos trabajos realizados en los campamentos de verano del COM.

9.1.4.1 Proyectos fin de carrera

A continuación, se describen los proyectos fin de carrera que se han desarrollado en la ETSIT y con los que se ha colaborado directamente para la validación de los resultados de la tesis doctoral:

- **Desarrollo de un prototipo de arquitectura EDA para el procesamiento y visualización de eventos de negocio en el ámbito bancario** (Montero, 2014). En este proyecto se desarrolló un prototipo de arquitectura EDA que se basa en la generación de eventos personales de negocio y en su explotación. También se realizaron distintos análisis de prestaciones del bróker de mensajería RabbitMQ.
- **Desarrollo de una aplicación de pago entre personas a través de redes sociales** (Cases, n.d.). En este proyecto se incluyen los detalles relacionados con la implementación de un servicio de pago directo entre personas a través de la red social Facebook que fue diseñado en el proyecto POSDATA.

Asimismo, la autora del proyecto ha colaborado en la validación de otras contribuciones a través de distintas becas, entre las que destacan: prueba de concepto

para la recuperación de datos públicos en servicios de redes sociales y el desarrollo de una infraestructura para el almacenamiento semántico.

9.1.4.2 Campamentos de verano del Center for Open Middleware

Durante los veranos de 2013 y 2014 se llevaron a cabo distintos campamentos de verano dentro del COM que permitieron a varios estudiantes de la ETSIT y la Facultad de Informática (FI) de la UPM familiarizarse con los trabajos de investigación del centro y llevar a cabo estudios de viabilidad de tecnologías relacionadas con los proyectos. En particular, en esta tesis doctoral se trabajó con tres alumnos de la ETSIT que llevaron a cabo los siguientes trabajos que contribuyen a la validación del enfoque global:

- Estudio preliminar sobre la viabilidad del procesador de eventos Esper, centrándose en aspectos de configuración de la herramienta, creación de módulos de sentencias y modificación de parámetros en tiempo de ejecución.
- Implementación de distintas funcionalidades en el entorno de desarrollo de Facebook. Concretamente, se realizaron las siguientes actividades: creación de perfiles de prueba, recuperación de datos personales, generación de publicaciones aleatorias, almacenamiento de información, etc.
- Estudio de distintas herramientas de visualización basadas en grafos para la representación de información en redes sociales.

9.2 Resultados obtenidos

Los resultados obtenidos en esta tesis doctoral se han protegido mediante dos patentes y han sido publicados en una revista con índice de impacto JCR y en distintos congresos internacionales y nacionales de relevancia. A continuación se resumen dichos trabajos y los premios obtenidos.

9.2.1 Protección intelectual e industrial a través de patentes

Fruto de la colaboración con la empresa Ericsson en el proyecto Segur@, se elaboraron distintas propuestas de ideas innovadoras patentables (*Invention Disclosures*) que tras ser analizadas y revisadas originaron su protección mediante dos patentes.

Dichas patentes están en explotación por parte de la empresa multinacional Ericsson, formando parte de su portfolio, en virtud del acuerdo de tramitación conjunta de patentes (*Patent Prosecution Management Agreement*) que se elaboró y firmó por parte de Ericsson y la Universidad Politécnica de Madrid en septiembre de 2010.

- **Method for Privacy Management in an Identity Network, Physical Entities and Computer Programs Therefor** (Monjas et al., 2010). En esta patente se describen los métodos relacionados con la gestión de la privacidad en una solución de gestión de identidad federada. La solicitud realizada ante la Organización Mundial de la Propiedad Intelectual (*World Intellectual Property Organization* – WIPO) se publicó el 14.10.2010

con número WO/2010/115460 y, tras su examen positivo, se encuentra en distintas fases dentro de oficinas nacionales de Europa, Japón, Australia, China y Estados Unidos. En este último, se concedió la patente el 12.8.2014 con referencia 13263669⁸⁹.

- **Method for Selectively Distributing Information in a Computer or Communications Network, and Physical Entities Therefor** (Monjas et al., 2012). Esta patente está relacionada con los trabajos de distribución de información en distintas redes sociales que sustenta el modelo de negocio de publicidad en redes sociales presentado. Se publicó el 05.01.2012 con número WO/2012/000543 y se encuentra en distintas fases de las oficinas nacionales de Europa, Estados Unidos e India⁹⁰.

9.2.2 Diseminación de resultados

Los resultados obtenidos en los distintos proyectos y con ello, las contribuciones de la tesis doctoral también se han descrito y enviado para su publicación en distintos congresos y revistas internacionales. A continuación, se describen los resultados obtenidos.

9.2.2.1 Artículos en revista

Confirmada la publicación del siguiente artículo en la revista IEEE América Latina con índice de impacto en el JCR SCI. Versión extendida e invitada para su publicación dada la selección entre los cinco mejores del VII Congreso Iberoamericano de Telemática (CITA2015):

- *A Personal Data Framework for Distributed User Modelling in New Service Ecosystems* (San Miguel, et al., 2015e).

9.2.2.2 Artículos en capítulos de libro

El siguiente artículo se incluyó como capítulo de libro:

- *Integrating Business Information Streams in a Core Banking Architecture: A Practical Experience* (San Miguel, et al., 2015d). Incluido como capítulo del libro *Enterprise Information Systems: 16th International Conference on Enterprise Information Systems (ICEIS 2014)*. Versión extendida tras ser seleccionado como mejor artículo del congreso.

9.2.2.3 Artículos en congresos internacionales y capítulos de libro

Los siguientes artículos han sido presentados en distintos congresos internacionales e incluidos en las actas de éstos, algunos de ellos fueron seleccionados entre los mejores de los congresos tras su presentación.

Se listan los artículos por orden cronológico de más a menos reciente y se señalan los aspectos más relevantes.

⁸⁹ WIPO, Patentscope WO2010115460,
<https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2010115460>

⁹⁰ WIPO, Patentscope WO2012000543,
<https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2012000543>

- *Extracting Knowledge through Friendsourcing: Application to new Financial Services* (San Miguel, et al., 2015d). Presentado en VII Congreso Iberoamericano de Telemática (CITA2015) y **seleccionado entre los cinco mejores del congreso**.
- *Approaching Unbanked People: A Person-to-Person Payment Application on Facebook* (San Miguel, et al., 2015d). Presentado en VII Congreso Iberoamericano de Telemática (CITA2015).
- *A Personal Data Framework for Exchanging Knowledge about Users in New Financial Services* (San Miguel, et al., 2015a). Presentado en *1st International Workshop on Personalization & Recommender Systems in Financial Services (FINREC)*.
- *Creating and Modelling Personal Socio-Economic Networks in On-Line Banking* (San Miguel, et al., 2014c). Incluido en el libro *Web Information Systems Engineering* al ser **seleccionado entre los mejores de los presentados** en *7th International Workshop on Context-Awareness and Personalization in Cloud and Service Computing (PCS 2014)*.
- *Evolving a Core Banking Enterprise Architecture: Leveraging Business Events Exploitation* (San Miguel, et al., 2014b). Presentado en *16th International Conference on Enterprise Information Systems (ICEIS 2014)* y seleccionado como mejor artículo del congreso en el área de arquitectura empresarial **[Best paper award]. Computer Science Conference Rank C [15% acceptance rate]**.
- *Personal Data Usage and Management* (San Miguel, et al., 2014a). Presentado en *Doctoral Consortium on Enterprise Information Systems (DCEIS 2014)*.
- *A Proposal for Enhancing the Motivation in Students of Computer Programming* (San Miguel, et al., 2012). Presentado en *5th International Conference of Education, Research and Innovation (ICERI 2012)*.
- *Modelado Semántico y Centrado en el Usuario de Servicios Adaptados al Contexto de Uso* (Yelmo, et al., 2011a). Artículo presentado en el VI Congreso Ibero-americano de Telemática (CITA2011).
- *Semantic Characterization of Context of Use and Contents for User-Centric Media Retrieval* (Yelmo, et al., 2010). Incluido en el libro *User Centric Media* al ser **seleccionado entre los mejores artículos** de *Second International ICST Conference, UCMedia 2010*.
- *Self-Service Privacy: User-Centric Privacy for Network-Centric Identity* (del Álamo, et al., 2010). Presentado en *4th IFIP WG 11.11 International Conference* e incluido como capítulo en el libro *Trust Management IV*.
- *Prosumers and accessibility: How to Ensure a Productive Interaction* (Martin, et al., 2009). Presentado en *2009 International Cross-Disciplinary Conference on Web Accessibility (W4A)*. **[John M Slatin Award for Best Communication Paper]**.

9.2.2.4 Congresos nacionales y otras publicaciones

Además de lo anterior se han publicado otros artículos de ámbito nacional entre los que destacan:

- *El Rol Central del Usuario en los Servicios Telemáticos* (Yelmo, et al., 2011b). White paper del proyecto Buscamedia.
- *Análisis del potencial de las comunicaciones móviles para la mejora de los servicios proporcionados por la Administración Pública a las personas con discapacidad* (Yelmo, et al., 2007b). Presentado en X Jornadas sobre Tecnologías de la Información para la Modernización de las Administraciones Públicas
- *Contribución a las metodologías de evaluación de accesibilidad de contenidos para la Web Móvil* (Yelmo, et al., 2007a). Presentado en el II Congreso Español de Informática (MWeb '07 Jornadas Sobre Web Móvil).
- Herramienta de evaluación de accesibilidad de contenidos y aplicaciones para terminales móviles (Yelmo, et al., 2006). Presentado en XVI Jornadas Telecom I+D. Madrid, 2006.

9.2.3 Premios otorgados

Los resultados obtenidos han sido galardonados con premios de distintas instituciones, así como en las conferencias donde han sido presentados. Los premios son:

- **Seleccionado entre los cinco mejores artículos** del VII Congreso Iberoamericano de Telemática (CITA2015) la publicación (San Miguel, et al., 2015c).
- **Seleccionado entre los mejores artículos** del *7th International Workshop on Context-Awareness and Personalization in Cloud and Service Computing (PCS 2014)* la publicación (San Miguel, et al., 2014c).
- **Mejor artículo** en el área de arquitectura empresarial del *16th International Conference on Enterprise Information Systems (ICEIS 2014)* para (San Miguel, et al., 2014b) con un ratio de aceptación del 15%.
- **Patentable Invention Awards** otorgados por Ericsson España como reconocimiento de la labor inventiva en los trabajos relacionados con el proyecto Segur@ y las patentes (Monjas, et al., 2010) y (Monjas, et al., 2012).
- **Accésit 1ª fase de la VII Competición de Creación de Empresas UPM** por la propuesta en 2010 de un modelo de negocio basado en soluciones para el pequeño comercio.
- **Seleccionado entre los mejores artículos** del congreso *Second International ICST Conference* la publicación (Yelmo, et al., 2010).

- ***John M Slatin Award for Best Communication Paper*** para el artículo (Martín, et al., 2009) presentado en *2009 International Cross-Disciplinary Conference on Web Accessibility (W4A)*.

9.3 Resumen del capítulo

En este capítulo se han recopilado todos los resultados y méritos de investigación derivados de la realización de la tesis doctoral. Éstos se han obtenido principalmente dada la participación en cuatros proyectos de investigación de ámbito nacional, dos de ellos de distintas convocatorias CENIT y otros dos del COM, una iniciativa del Banco Santander en colaboración con la UPM. Además, se han descrito los trabajos realizados dentro del contexto académico que han permitido validar el enfoque global de la tesis doctoral con distintos estudios de viabilidad de tecnologías y la implementación de pruebas de concepto de las ideas surgidas en los proyectos.

Los trabajos realizados han derivado en la obtención de dos patentes, un artículo en una revista indexada en el JCR SCI y múltiples artículos en congresos internacionales de relevancia. Además, se han galardonado distintos trabajos por la calidad de los mismos.

Capítulo 10.

Conclusiones

Los datos personales se han convertido en una fuente de generación de valor tanto para los proveedores de servicios telemáticos como para los usuarios de éstos. En este sentido, para los primeros, los datos que obtienen de sus usuarios suponen un nuevo activo económico que permite configurar sus modelos de negocio. Por otra parte, para los usuarios, la utilización y el procesamiento de sus datos les permite acceder a servicios telemáticos basados en ofrecer una experiencia de uso adaptada a sus preferencias y necesidades personales, lo cual se percibe como un valor añadido sobre los beneficios de dichos servicios como, por ejemplo, el acceso ubicuo, la disponibilidad permanente, la independencia del dispositivo utilizado, la multimodalidad y habitualmente, el uso gratuito.

A pesar de los valores y beneficios asociados a la utilización de los datos personales para los actores involucrados en la provisión de servicios telemáticos, dichos recursos se encuentran bajo la custodia de los proveedores de servicios y esto suscita dudas respecto a la ética, conducta y responsabilidades de sus modelos de negocio, puesto que una utilización inadecuada o inapropiada de los datos personales se traduce en un riesgo para la privacidad del usuario.

Ante este contexto, esta tesis doctoral ha propuesto un nuevo modelo de negocio que permite cubrir las debilidades y amenazas de los existentes para promover la creación, compartición y utilización de datos personales, y en particular, modelos de usuario, entre distintos proveedores que difícilmente pueden alcanzarse de manera unilateral. Al mismo tiempo, este modelo ofrece a los usuarios una propuesta de valor en la que pueden establecer su control en cuanto a la cesión y uso de éstos y obtener nuevos beneficios respecto a sus recursos personales.

Para lo anterior, se ha abordado el estudio de tres áreas de investigación que se originan en torno a los datos personales: modelado de usuario, gestión de identidad y privacidad, y se ha propuesto la combinación y el avance de éstas en una arquitectura de referencia que se basa en la recuperación de recursos externos, la generación de conocimiento de los usuarios, la

provisión de éste a terceras entidades y la incorporación de herramientas centradas en el usuario para la gestión de su privacidad.

En la siguiente sección se enumeran los objetivos que se establecieron al inicio de esta investigación y se resaltan las conclusiones más relevantes extraídas de su consecución. Seguidamente, en la sección 10.2 se incluyen las contribuciones originales de la tesis y para finalizar, en la sección 10.3 se presentan las líneas de investigación que pueden abordarse como trabajo futuro.

10.1 Consecución de los objetivos originales de la tesis

A continuación, se destacan los objetivos específicos definidos al inicio de la tesis, indicando en qué capítulos del documento se han abordado y los resultados y conclusiones más relevantes:

- **Identificar y analizar los principales modelos de negocio aplicables dentro del ecosistema de datos personales actual con el objeto de proponer un nuevo modelo en torno a la utilización e intercambio de datos personales.**

Este objetivo se ha abordado en los capítulos 3 y 4 del documento. En primer lugar, en el capítulo 3 se incluyó un estudio sobre los modelos de negocio más relevantes en el contexto de servicios telemáticos y en particular, se profundizó en aquellos que se basan en el uso y la compartición de datos personales de los usuarios. Seguidamente, en el capítulo 4 se presentó el posicionamiento de esta tesis doctoral respecto al estudio señalado, la descripción de un nuevo modelo de referencia y su particularización en dos ámbitos de aplicación.

Los resultados obtenidos ponen de manifiesto que los datos personales suponen un activo económico esencial para la mayoría de los proveedores de servicios actuales y que su aprovechamiento puede utilizarse para generar valor tanto para los propios proveedores como para sus usuarios. A pesar de que existen modelos de negocio como, por ejemplo, el basado en publicidad o el de los brókeres de datos personales, que actúan en beneficio propio, también existen nuevas oportunidades basadas en la generación de valor para los usuarios que permiten cubrir las debilidades y amenazas de los modelos actuales.

- **Investigar y analizar las áreas de investigación y tecnologías asociadas relacionadas con la obtención de conocimiento y la gestión de datos personales de los usuarios.**

Los resultados más relevantes de este objetivo se han incluido en el capítulo 3 donde se han descrito tres áreas de investigación principales: modelado de usuario, gestión de identidad y privacidad. Aquí, se ha puesto énfasis en las soluciones que permiten la aplicación de estas áreas en entornos distribuidos en los que participan distintos proveedores de servicios y usuarios. En este sentido, se ha detectado que existen algunas soluciones puntuales que externalizan el proceso de modelado de usuario, sin embargo éstas no abordan de forma global aspectos relacionados con la recuperación de datos, la gestión de la identidad o la gestión de la privacidad, lo cual ha dado pie a

proponer un nuevo proceso de modelado de usuario, cuyas etapas se describieron en los capítulos 6, 7 y 8.

- **Especificar y diseñar una arquitectura de referencia que posibilite la realización del modelo de negocio identificado. Esta arquitectura debe apoyarse en el abastecimiento de recursos externos, la generación de conocimiento de los usuarios, la provisión de éste a terceras entidades y la incorporación de herramientas centradas en el usuario para la gestión de su privacidad.**

Los resultados obtenidos de este objetivo se han incluido en el capítulo 5 del documento donde se ha establecido una especificación que no sólo considera los servicios ofrecidos a los actores principales, sino que también establece dos principios fundamentales que determinan el correcto funcionamiento del sistema: interoperabilidad y privacidad. Dicha especificación ha permitido diseñar una arquitectura de referencia que se ha detallado también en el capítulo 5, aunque los detalles y validaciones de sus módulos se han presentado en los capítulos posteriores 6, 7 y 8.

La arquitectura establecida es una solución ambiciosa que cubre distintos retos relacionados con la recolección, generación, recuperación, gestión de identidad y privacidad de datos personales y modelos de usuario. Además, involucra a distintos actores externos (los usuarios, fuentes de datos y consumidores de modelos de usuario) y pone de manifiesto que la privacidad puede abordarse como una funcionalidad más del sistema puesto que tiene valor de negocio propio.

- **Identificar y caracterizar las fuentes de datos personales con potencial de abastecer de los recursos necesarios a la arquitectura de referencia establecida en el punto anterior.**

Este objetivo se ha llevado a cabo en el capítulo 6 del documento. Aquí, se han caracterizado hasta doce fuentes de datos conforme a su utilidad dentro del proceso de modelado de usuario distribuido, los recursos que custodian, los tipos de acceso a éstos y los mecanismos asociados para su recuperación.

De lo anterior se desprende que actualmente existen muchas posibilidades para la obtención de datos personales de los usuarios, no obstante para llevar a cabo su recuperación se establece una relación de dependencia con la fuente de datos considerada y en particular, con los servicios que ésta incluya para la recuperación de sus recursos. En este sentido, se ha profundizado en aquellos mecanismos que permiten la recuperación de recursos a través de tecnologías para la delegación de autorización en las que el usuario es el protagonista para la recuperación de sus datos personales.

- **Determinar distintos modelos para la generación de conocimiento de los usuarios a partir de recursos distribuidos y heterogéneos. En particular, se deberán explorar soluciones basadas en tecnologías semánticas.**

En el capítulo 7 de este documento se han presentado los trabajos relacionados con este objetivo y así, se han descrito tres patrones de generación principales que permiten la producción de modelos de usuario con distintas características a partir de recursos heterogéneos y distribuidos.

Los resultados obtenidos en la consecución de este objetivo ponen de manifiesto que el proceso de generación de modelos de usuario depende principalmente de dónde se encuentren los recursos necesarios para la producción y de la temporalidad y la complejidad de los modelos de usuario requeridos por las entidades que vayan a hacer uso de éstos. De esta forma, se han establecido patrones de generación que se basan en soluciones interoperables para la centralización de recursos heterogéneos que permiten la producción de modelos de usuario complejos y otros patrones que se apoyan en mecanismos para la generación de modelos dinámicos y con validez a corto plazo.

Cabe señalar en este punto que, tal y como se expuso en el capítulo 5, se requieren soluciones interoperables que permitan establecer vocabularios y terminologías comunes y formatos de intercambio homogéneos de datos, entre otros, que promuevan el uso e intercambio de datos y con ello, la generación de modelos de usuario. En este sentido, en los capítulos 3, 6 y 7 se han señalado distintas iniciativas y soluciones puntuales establecidas en ámbitos de aplicación muy concretos pero que no abordan todas las perspectivas de interoperabilidad consideradas.

- **Establecer una infraestructura de gestión de identidad y privacidad centrada en el usuario que permita habilitar la gestión del flujo de sus datos personales para establecer el control en el uso y la cesión de éstos.**

Este objetivo se ha abordado en el capítulo 8 del documento donde se ha establecido una infraestructura para la gestión de la identidad que permite habilitar los flujos de recuperación de datos y de recuperación de modelos de usuario que se establecen en la arquitectura de referencia descrita en el capítulo 5. Además, lo anterior ha posibilitado el establecimiento de una infraestructura de gestión de privacidad donde el usuario puede controlar los flujos habilitados para establecer sus preferencias respecto al uso y cesión de sus recursos personales.

La conclusión más relevante de los resultados del capítulo 8, así como el estado del arte relacionado en el capítulo 3, es que la privacidad debe entenderse como un complemento esencial de la gestión de la identidad para ofrecer a los usuarios la confianza respecto al uso e intercambio de sus datos personales.

- **Validar los resultados obtenidos a través de distintos prototipos y contribuciones en el ámbito de proyectos de investigación industrial aplicada.**

En cada uno de los capítulos que contiene contribuciones originales de la tesis (capítulos 4, 5, 6, 7 y 8) se ha incluido una sección específica que describe las validaciones realizadas. Éstas han venido definidas por los proyectos de investigación en los que ha participado la autora de esta tesis y se han descrito en el capítulo 9.

Principalmente, se han considerado validaciones relacionadas con servicios de redes sociales y nuevos servicios financieros y en este sentido, se han desarrollado distintos prototipos para la recuperación de datos personales, para la generación de modelos de usuario, para la aplicación de éstos y para permitir al usuario la gestión de sus identidades digitales y el establecimiento de políticas de privacidad.

- **Diseminar los resultados obtenidos en revistas y congresos internacionales relevantes de temática relacionada, así como adoptar las medidas necesarias para proteger su propiedad intelectual e industrial a través de patentes.**

En el capítulo 9 se ha incluido una sección específica con los resultados derivados de esta investigación. Cabe destacar que se han obtenido dos patentes en explotación, un artículo en una revista indexada en el JCR SCI y doce publicaciones de congresos internacionales o como capítulos de libros, algunos de éstos han sido galardonados con premios.

10.2 Contribuciones de la tesis doctoral

Como resultado de las actividades de investigación llevadas a cabo se ha elaborado un conjunto de contribuciones originales que se han ido detallando en la sección final de los capítulos correspondientes. A continuación, se agrupan las contribuciones originales:

- Respecto a los **nuevos modelos de negocio** para el ecosistema de datos personales actual presentados en el capítulo 4, se ha contribuido con:
 - Análisis crítico de los distintos modelos de negocio que existen actualmente respecto al uso y la compartición de datos personales con terceras entidades para detectar nuevas oportunidades de negocio.
 - Propuesta de un nuevo modelo de negocio de referencia que aprovecha las oportunidades detectadas, fomentando como propuestas de valor de éste las debilidades de los modelos de negocio actuales.
 - Particularización del modelo de negocio en el sector publicitario, lo cual incluye un análisis específico del contexto publicitario y la detección de huecos existentes para la propuesta de un nuevo modelo de negocio basado en la intermediación de información respetando la privacidad del usuario.
 - Particularización del modelo de negocio en el sector financiero, lo cual implica un estudio del sector, detectando las últimas tendencias y la evolución del mercado y con ello, la propuesta de un nuevo modelo de negocio que permite posicionarse a los proveedores tradicionales para la promoción de nuevos servicios financieros.
- Dentro del capítulo 5 se ha presentado la especificación y diseño de **una arquitectura genérica de referencia** que permite la recuperación de recursos custodiados por distintas entidades (fuentes de datos) para generar modelos de usuario enriquecidos

que cubren las necesidades específicas de terceras entidades, considerando la participación del usuario y el control sobre sus recursos personales (datos y modelos de usuario).

- En el capítulo 6 se ha profundizado en el proceso de **recuperación de recursos** custodiados por fuentes de datos externas y las contribuciones originales de la tesis son:
 - La identificación y caracterización de las fuentes de datos externas con potencial de abastecer al marco de datos personales de los recursos necesarios. Las contribuciones originales a este respecto son: la consideración del usuario como fuente de datos y la identificación de un nuevo tipo de datos personales (eventos personales de negocio) que puede originarse dentro de una arquitectura empresarial.
 - Propuesta de un modelo global para la recuperación de recursos heterogéneos distribuidos en distintas fuentes de datos.
 - Diseño y desarrollo de un nuevo servicio de pago personal que a su vez ha permitido validar la recuperación de recursos personales en distintos proveedores de servicios de redes sociales y servicios de mensajería.
 - Definición de un nuevo tipo de datos personales identificado en el contexto de esta investigación que ha sido denominado como eventos personales de negocio.
- Respecto al proceso de **generación de modelos de usuario** que se ha abordado en el capítulo 7, las contribuciones son:
 - Establecimiento de tres patrones principales que permiten llevar a cabo la generación de modelos de usuario considerando recursos heterogéneos y distribuidos.
 - Definición de una ontología que permite representar características socio-económicas de los usuarios y obtener modelos de usuario complejos a partir de ésta.
 - Diseño y desarrollo de un nuevo servicio financiero que ofrece a sus usuarios la posibilidad de recibir una cantidad determinada de dinero a través de micropréstamos de sus amigos. Para ello, se apoya en el análisis de modelos de usuario socio-económicos para generar un conjunto de prestadores que satisfacen un grado de confianza, liquidez y solvencia dados.
 - Identificación y validación de distintos mecanismos de generación y recuperación de conocimiento de los usuarios basados en un nuevo tipo de datos personales (eventos personales de negocio) establecido en esta investigación dentro de una arquitectura empresarial, en concreto, del ámbito bancario.

- Por último, en el capítulo 8 se han descrito las **infraestructuras de gestión de identidad y gestión de privacidad** que permiten la realización de los procesos anteriores y las contribuciones obtenidas son:
 - Análisis de los flujos de datos que se establecen en torno a la arquitectura establecida e identificación de aquellos sobre los cuales el sistema tiene capacidad para determinar su ejecución y habilitar su gestión al usuario.
 - Propuesta de una infraestructura para la gestión de identidad centrada en el usuario que permite a éste la administración de modelos de gestión centralizados y federados establecidos entre distintos proveedores de servicios.
 - Propuesta de una infraestructura para la gestión de la privacidad que se basa en el establecimiento de políticas de privacidad conforme a los intereses y preferencias del usuario y distintas funcionalidades que guían al usuario sobre la toma de decisiones en el uso y compartición de sus recursos personales.

10.3 Trabajo futuro

En esta tesis doctoral se ha abordado el estudio, la combinación y el avance de tres áreas de investigación principales que se originan en torno a los datos personales de los usuarios: modelado de usuario, gestión de identidad y privacidad, y su aplicación en un ecosistema de intercambio de datos centrado en el usuario. A lo largo del desarrollo de esta investigación se han detectado algunos aspectos de dichas áreas cuyo estudio puede resultar de especial interés pero que se no se ha abordado por encontrarse fuera de los objetivos establecidos o por no contar con los recursos necesarios. Asimismo, existen otras áreas de investigación relacionadas con el tratamiento de datos que pueden ser de utilidad y que junto con lo anterior pueden servir de punto de partida para futuras líneas de investigación que den continuidad a los resultados obtenidos en esta tesis doctoral.

En primer lugar, y tal y como se destacó en el Capítulo 3 de Estado del arte respecto a la privacidad, lo que se ha dado en llamar Ingeniería de la privacidad es todavía una disciplina incipiente que requiere de trabajos y soluciones que permitan materializar tecnológicamente los principios establecidos por distintos gobiernos y organizaciones como, por ejemplo, los de la OECD (2013) o los de la Privacidad desde el diseño (Cavoukian, 2011). En este sentido, se propone el estudio e investigación de nuevas soluciones que contribuyan a la protección de la privacidad de los usuarios a través de nuevos mecanismos que consideren, por ejemplo, la anonimización (disociación) de datos, la minimización en la recolección y la generación de recursos personales y, en general, aquellos basados en los principios mencionados.

Por otra parte, la complejidad y dimensión de los datos personales, así como las bases de datos asociadas, han originado áreas de investigación como, por ejemplo, la de Datos masivos (*Big data*) y la de Ciencia de los Datos (*Data Science*), que pueden resultar de interés para considerar aspectos de escalabilidad de las soluciones propuestas o el establecimiento de

nuevos patrones de generación de modelos de usuario basados en la estadística o matemáticas aplicadas.

Referencias

- Abril, P. S. y Pizarro, E., 2014. La intimidación europea frente a la privacidad americana – una visión comparativa del derecho al olvido. *Indret: Revista para el Análisis del Derecho*, 1(2014).
- Allee, V., 2000. Reconfiguring the Value Network. *Journal of Business Strategy*, 21(4), pp. 36-39.
- Almeshekah, M. H., 2010. Economics of Identity Management Systems – Towards an Economically Adaptive User-Centric IdMS. MSc in Information Security. University of London. Disponible en: <https://www.meshekah.com/research/publications_files/ms_thesis_identity_manageme.pdf> [Consulta: septiembre 2015].
- Álvarez, M., 2014. Apertura y reutilización de datos públicos. Colección Govern Obert. Generalitat de Catalunya. Disponible en: <<http://transparencia.gencat.cat/es/menu/colleccio-govern-obert/>> [Consulta: mayo 2015].
- Anderson, C., 2006. *The Long Tail: Why the Future of Business is Selling Less of More*. New York: Hyperion.
- Arango Serna, M. D., Londoño Salazar, J. E. y Zapata Cortés, J. A., 2010. Arquitectura empresarial – una visión general. *Revista Ingenierías Universidad de Medellín*, 9 (16), pp. 101-111.
- Asif, M., 2014. Personalization of Mobile Services. Doctoral Thesis. Norwegian University of Science and Technology (NTNU). Disponible en: <https://www.idi.ntnu.no/research/doctor_theses/muhamma.pdf> [Consulta: junio 2015].
- Augure, 2014. Informe Augure sobre el Estatus del Marketing de Influencers. Disponible en: <<http://www.augure.com/es/blog/influencer-marketing-estatus-20140220>> [Consulta: junio 2015].
- Barla, M., 2010. Towards Social-based User Modeling and Personalization. Dissertation Thesis. Slovak University of Technology in Bratislava. Disponible en: <<http://acmbulletin.fiit.stuba.sk/theses/barla-thesis.pdf>> [Consulta: octubre 2015].
- Barri, I., Loilier, T., van Rijn, M., Stolk, A. and Vasiliadis, H., 2014. Open innovation in the financial services sector – Why and how to take action. Blue paper GFT Technologies.
- Basole, R. C. and Rouse, W. B., 2008. Complexity of service value networks: Conceptualization and empirical investigation. *Journal IBM Systems*, 47 (1), pp. 53-70.
- Bernstein, M. S., Tan, D., Smith, G., Czerwinski, M. and E. Horvitz, 2008. Personalization via Friendsourcing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(2), pp. 6:1-6:28.

- Bertino, E. and Takahashi, K., 2011. Identity Management. Concepts, Technologies and Systems. Norwood: Artech House.
- Bertocci, V., Serack, G. and Baker, C., 2007. Understanding Windows Cardspace: An Introduction to the Concepts and Challenges of Digital Identities. Addison-Wesley Professional.
- Bhargav-Spantzel, A., Camenisch, J., Gross, T. and Sommer, D., 2007. User Centricity: a Taxonomy and Open Issues. Journal of Computer Security. IOS Press, 15(5), pp. 493-527.
- Bizer, C., Heath, T. and Berners-Lee, T., 2009. Linked Data – The story so far. International Journal on Semantic Web and Information Systems, 5(3), pp. 1-22.
- Brickley, D. and Miller, L., 2014. Foaf Vocabulary Specification 0.99. Paddington Edition. Disponible en: <<http://xmlns.com/foaf/spec/20140114.html>> [Consulta: junio 2015].
- Brusilovsky, P., 1996. Methods and Techniques of Adaptive Hypermedia. User Model and User-Adapted Interaction, 6(2-3), pp. 87-129.
- Brusilovsky, P. and Millán, E., 2007. User Models for Adaptive Hypermedia and Adaptive Educational Systems. In: P. Brusilovsky, A. Kobsa and W. Neidl, eds. 2007. The Adaptive Web: Methods and Strategies of Web Personalization. New York: Springer-Verlag. Lecture Notes in Computer Science, 4321, pp. 3-53.
- Buscamedia, 2015. CENIT Buscamedia. [En línea] Disponible en: <<http://www.cenitbuscamedia.es>> [Consulta: junio 2015].
- Busch, W. and Moreno, J. P., 2014. Banks' New Competitors: Starbucks, Google, and Alibaba. Harvard Business Review. Disponible en: <<https://hbr.org/2014/02/banks-new-competitors-starbucks-google-and-alibaba/>> [Consulta: junio 2015].
- Cabrera, J., 2013. Diseño de un crawler modular para buscadores verticales. Proyecto final de carrera. Universidad Politécnica de Cataluña. Disponible en: <<http://upcommons.upc.edu/handle/2099.1/18140>> [Consulta: julio 2015].
- Cao, Y. and Yang, L., 2010. A survey of Identity Management Technology. IEEE International Conference on Information Theory and Information Security (ICITIS). Beijing, China, 17-19 December 2010. IEEE, pp. 287-293.
- Carbonell, J. R., 1970. AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction. IEEE Transactions on Man-Machine Systems, 11(4) pp. 190-202.
- Carlsson, J., 2010. An Assessment of Social Media Business Models and Strategic Implications for Future Implementation. Diploma in Advanced Strategy. University of Oxford (Saïd) Business School.
- Carmagnola, F., Cena, F. and Gena, C., 2011. User model interoperability: a survey. User Modeling and User-Adapted Interaction, 21(3), pp. 285-331.

- Carrión Delgado, M. G., 2012. Diseño y desarrollo de un modelo computacional para la representación del conocimiento en el dominio de la cooperación judicial en materia penal. Tesis doctoral. Universidad Nacional de Educación a Distancia. Disponible en: <<http://espacio.uned.es/fez/eserv/tesisuned:Filologia-Mgcarrion/Documento.pdf>> [Consulta: octubre 2015].
- Cases, N. N., n.d. Desarrollo de una aplicación de pago entre personas a través de redes sociales. Proyecto fin de carrera. Universidad Politécnica de Madrid.
- Castells, M., 2006. Informacionalismo, redes y sociedad red: una propuesta teórica. La sociedad red : una visión global. Alianza Editorial, pp. 27-75.
- Castillo, C, 2004. Effective Web Crawling. Ph.D. in Computer Science. Universidad de Chile. Disponible en: <http://chato.cl/papers/crawling_thesis/effective_web_crawling.pdf> [Consulta: julio 2015].
- Cavoukian, A., 2011. Privacy by Design – 7 Foundational Principles. [En línea] Disponible en: <<https://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/>> [Consulta: julio 2015].
- Chaia, A., Dalal, A., Goland, T, Gonzalez, M. J., Morduch, J. and Schiff R., 2009. Half the World is Unbanked. Financial Access Initiative Framing Note. Disponible en: <http://mckinseyonsociety.com/downloads/reports/Economic-Development/Half_the_world_is_unbanked.pdf> [Consulta: septiembre 2015].
- Chesbrough, H. W., 2003. Open Innovation: The New Imperative for Creating And Profiting from Technology. Boston: Harvard Business School Press.
- Comisión Europea, 2008. Financial Services Provision and Prevention of Financial Exclusion. Directorate-General for Employment, Social Affairs and Equal Opportunities Inclusion, Social Policy Aspects of Migration, Streamlining of Social Policies. Disponible en: <http://www.libertysecurity.org/IMG/pdf_financial_exclusion_summary_en.pdf> [Consulta: noviembre 2015]
- Comisión Europea, 2011. Attitudes on Data Protection and Electronic Identity in the European Union. Special Eurobarometer 359. Conducted by TNS Opinion & Social at the request of Directorate-General Justice, Information Society & Media and Joint Research Centre. Disponible en: <http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf> [Consulta: noviembre 2015]
- Comisión Europea, 2014. Green paper on mobile Health ("mHealth"). [En línea] Disponible en: <<http://ec.europa.eu/digital-agenda/en/news/green-paper-mobile-health-mhealth>> [Consulta: noviembre 2015].
- Comisión Europea, 2015. Data Protection. Special Eurobarometer 431. Conducted by TNS Opinion & Social at the request of Directorate-General for Justice and Consumers. Disponible en: <http://ec.europa.eu/public_opinion/archives/ebs/ebs_431_en.pdf> [Consulta: noviembre 2015]

- Comisión Nacional de los Mercados y la Competencia (CNMC), 2014. Efectos de los nuevos modelos económicos sobre el mercado y la regulación. Departamento de Promoción de la Competencia. Disponible en: http://www.cnmc.es/Portals/0/Ficheros/Promocion/Informes_y_Estudios_Sectoriales/2014/EconomiaColaborativa/20141124%20Consulta%20Doc2%20%20Efectos%20nuevos%20modelos%20FINAL.pdf [Consulta: junio 2015].
- Davoli, A., 2015. La protección de los datos personales. Fichas técnicas sobre la Unión Europea – 2015. [En línea] Disponible: http://www.europarl.europa.eu/ftu/pdf/es/FTU_5.12.8.pdf [Consulta: noviembre 2015].
- Del Álamo, J. M., 2009. Contribution to Operation Support Systems and Service Management Architectures for User-Centric Telecommunications Services over Next Generation Networks. Tesis doctoral. Universidad Politécnica de Madrid.
- Del Álamo, J.M., Monjas, M.A., Yelmo, J.C., San Miguel, B., Trapero, R. and Fernández, A.M., 2010. Self-Service Privacy: User-Centric Privacy for Network-Centric Identity. In: M. Nishigaki et al. (eds.), Trust Management IV: 4th IFIP WG 11.11 International Conference, IFIPTM 2010, Morioka, Japan, June 16-18, 2010. Proceedings. Springer Berlin Heidelberg, pp. 17-31.
- Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Diario Oficial nº L 281 de 23/11/1995 p.0031 - 0050. Disponible en: <http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex:31995L0046> [Consulta: julio 2015].
- DOMO, 2015. Data Never Sleeps 3.0. Dataviz. [En línea] Disponible en: <https://www.domo.com/learn/data-never-sleeps-3-0> [Consulta: noviembre 2015].
- Eclipse Foundation, 2009. Higgins – Personal Data Service. [En línea] Disponible en: <http://www.eclipse.org/higgins/> [Consulta: julio 2015].
- Etzion, O. and Niblett, P., 2010. Event Processing in Action (1st ed.). Greenwich: Manning Publications Co.
- Eugster, P., Felver, P., Guerraoui, R. and Kermarrec, A., 2003. The Many Faces of Publish/Subscribe. Journal ACM Computing Surveys (CSUR), 35(2), pp.114-131.
- European Union Agency for Network and Information Security (ENISA), 2014. Privacy and Data Protection by Design – from policy to engineering. Disponible en: <http://arxiv.org/ftp/arxiv/papers/1501/1501.03726.pdf> [Consulta: julio 2015].
- Facebook, 2015. Facebook para empresas. Anuncios sencillos y efectivos. [En línea] Disponible en: <https://www.facebook.com/business/products/ads> [Consulta: junio 2015].
- Federal Trade Commission (FTC), 1998. Privacy Online: A Report to Congress. Washington: Federal Trade Commission. Disponible en: <https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report->

[congress/priv-23a.pdfhttps://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014](https://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014)> [Consulta: julio 2015].

Federal Trade Commission (FTC), 2014. Data Brokers: A call for Transparency and Accountability. Washington: Federal Trade Commission. Disponible en: <<https://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014>> [Consulta: mayo 2015].

Federal Trade Commission (FTC), 2015. Privacy and Security. Business Center Resources. [En línea] Disponible en: <<https://www.ftc.gov/tips-advice/business-center/privacy-and-securityhttps://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014>> [Consulta: noviembre 2015].

Fielding, R. T., 2000. Architectural Styles and the Design of Network-based Software Architectures. Ph.D. dissertation. University of California. Disponible en: <https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf> [Consulta: septiembre 2015].

Flores Vitelli, I. J., 2011. Introducción al Razonamiento Sobre Ontologías. Lecturas en Ciencias de la Computación. Universidad Central de Venezuela.

Frías-Martínez, E., Magoulas, G., Chen, S. and Macredie, R., 2005. Modeling Human Behavior in User-Adaptive Systems: Recent Advances Using Soft Computing Technique. Expert Systems with Applications, 29(2), pp. 320-329.

Fröschl, C., 2005. User Modeling and User Profiling in Adaptive E-learning Systems. Master's Thesis. Graz University of Technology. Disponible en: <<http://www.iicm.tugraz.at/thesis/cfroeschl.pdf>> [Consulta: junio 2015].

Fundación Telefónica, 2011. Real time web: una nueva conciencia global. Colección Fundación Telefónica. Barcelona: Editorial Ariel.

Future of Identity in the Information Society (FIDIS), 2005. D2.3 Models. Final Deliverable. Disponible en: <<http://www.fidis.net/resources/fidis-deliverables/identity-of-identity/#c1753>> [Consulta: septiembre 2015].

Gach, R. and Gotsch, M., 2014. The Rise of Fintech. New York's Opportunity for Tech Leadership. Informe de Accenture y PARTNERSHIP FUND for New York City. Disponible en: <<http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Rise-of-Fintech-New-York.pdf>> [Consulta: junio 2015].

Gafni, R. and Nissim, D., 2014. To social login or not login? – Exploring factors affecting the decision. Issues in Informing Science and Information Technology, 11, pp. 57-72. Disponible en: <<http://iisit.org/Vol11/IISITv11p057-072Gafni0462.pdf>> [Consulta: noviembre 2015].

Gantz J. and Reinsel, D., 2012. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. IDC iView, IDC Analyze the Future. Disponible en: <<https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>> [Consulta: noviembre 2015].

- Gat, I. and Succi, G., 2013. A Survey of the API Economy. Agile Product & Project Management Executive Update, 14(6).
- Gauch, S., Speretta, M., Chandramouli, A. and Micarelli, A., 2007. User profiles for personalized information access. In: P. Brusilovsky, A. Kobsa and W. Neidl, eds. 2007. The Adaptive Web: Methods and Strategies of Web Personalization. New York: Springer-Verlag. Lecture Notes in Computer Science, 4321, pp. 54-89.
- Gigya, 2015. The Landscape of Customer Identity: Facebook Dominates, Payment Providers on the Rise. [En línea] Disponible en: < <http://www.gigya.com/blog/the-landscape-of-customer-identity-q2-2015/>> [Consulta: noviembre 2015].
- Goel, V., 2015. Facebook Announces a Payments Feature for Its Messenger App. The New York Times, Technology, March 17.
- Golbeck, J., Parsia, B. and Hendler, J., 2003. Trust Networks on the Semantic Web. In: M. Klusch, A. Omicini, S. Ossowski and H. Laamanen, eds. 2003. Cooperative Information Agents VII: Proceedings 7th International Workshop, CIA 2003, Helsinki, Finland, August 27-29, 2003. Springer Berlin Heidelberg. Lecture Notes in Computer Science, 2782, pp. 238-249.
- Goldberg, I., 2003. Privacy-Enhancing Technologies for the Internet, II: Five Years Later. In: R. Dingledine and P. Syverson, eds. 2003. Privacy Enhancing Technologies: Second International Workshop, PET 2002 San Francisco, CA, USA, April 14-15, 2002 Revised Papers. Springer Berlin Heidelberg. Lecture Notes in Computer Science, 2482, pp. 1-12.
- Gonzalez-Carrasco, I., Colomo-Palacios, Ro, Lopez-Cuadrado, J. L. García-Crespo, A. and Ruiz-Mezcua, B., 2012. PB-ADVISOR: A private banking multi-investment portfolio advisor. Information Sciences, 206, pp. 63-82.
- Google, 2015. Download your data. [En línea] Disponible en: <<https://support.google.com/accounts/answer/3024190?hl=en>> [Consulta: junio 2015].
- Gouriten, G. and Senellart, P., 2013. API Blender: A Uniform Interface to Social Platform APIs. In: WWW 2012 Developer Track. Disponible en: <<http://arxiv.org/abs/1301.2086>> [Consulta: junio 2015].
- Guarino, N., 1998. Formal ontology and information systems. Proceedings of the First International Conference (FOIS'98). Trento, Italy, 6-8 June 1998. Amsterdam: IOS Press, pp. 3-15.
- Halim, K., 2011. Personal Data List in Mind Map form. Identity Woman. [En línea] Disponible en: <<http://www.identitywoman.net/personal-data-list-in-mind-map-form>> [Consulta: junio 2015].
- Heckmann, D., 2005. Ubiquitous User Modeling. Ph.D. dissertation. Universitat des Saarlandes. Disponible en: <<http://d-nb.info/978586085/34>> [Consulta: septiembre 2015].

- Heurix, J., Zimmermann, P., Neubauer, T. and Fenz, S., 2015. A Taxonomy for Privacy Enhancing Technologies. *Computers & Security*, 53(September 2015), pp. 1-17.
- IHS Electronics & Media, 2013. Wearable Technology – Market Assessment. An IHS Whitepaper. Disponible en: <<http://cdn2.hubspot.net/hub/396065/file-2568104498-pdf>> [Consulta: septiembre 2015].
- Information Card Foundation, 2011. Quick Overview. [En línea] Disponible en: <<http://informationcard.net/quick-overview>> [Consulta: julio 2015].
- Instituto de Ingeniería Eléctrica y Electrónica (IEEE), 2010. Standards Glossary. [En línea] Disponible en: <https://www.ieee.org/education_careers/education/standards/standards_glossary.html> [Consulta: julio 2015].
- Instituto Nacional de Estadística (INE), 2014. Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares – Año 2014. Notas de prensa. Disponible en: <<http://www.ine.es/prensa/np864.pdf>> [Consulta: septiembre 2015].
- International Standards Office (ISO), 2011. ISO/IEC 29100:2011 Information technology – Security techniques - Privacy framework.
- International Telecommunication Union (ITU), 2009. Baseline capabilities for enhanced global identity management and interoperability. Recommendation ITU-T X.1250.
- International Telecommunication Union (ITU), 2010. Términos y definiciones de referencia para la gestión de la identidad. Recomendación ITU-T X.1252.
- Internet Engineering Task Force (IETF), 2012. The OAuth 2.0 Authorization Framework. Proposed Standard RFC 6749.
- Internet Engineering Task Force (IETF), 2014. The JavaScript Object Notation (JSON) Data Interchange Format. Proposed Standard RFC 7159.
- Ipsos MORI, 2013. Road to Inclusion: A Look at the Financial Underserved and Excluded across Europe. White paper for Mastercard Worldwide. Disponible en: <https://www.ipsos-mori.com/DownloadPublication/1616_marketing-ethnography-road-to-inclusion-mastercard-2013.pdf> [Consulta: octubre 2015].
- Jansen, B. J., Zhang, M. Sobel, K. and Chowdury, A., 2009. Twitter Power: Tweets as Electronic Word of Mouth. *Journal of the American Society for Information Science and Technology*, 60 (11), pp. 2169-2188.
- Jøssang, A. and Pope, S. 2005. User Centric Identity Management. Proceedings of AusCERT Conference 2005. Gold Coast, Australia, May 2005.
- Kabir, M.A., Han, J., Yu, J. and Colman, A., 2014. User-centric social context information management: an ontology-based approach and platform. *Personal and Ubiquitous Computing*, 18(5), pp. 1061-1083.

- Kay, J., 1994. Lies, damned lies and stereotypes: Pragmatic approximations of users. In: ACM 4th International Conference on User Modeling, pp. 175-184.
- Kim, G., Tao, W., Shin, N. and Kim, K. S., 2010. An empirical study of customers' perceptions of security and trust in e-payment systems. *Electronic Commerce Research and Applications*, 9 (1), pp. 84-95.
- Kim, Y. J., 2009. Access Control Service Oriented Architecture Security. [En línea] Disponible en: <<http://www.cse.wustl.edu/~jain/cse571-09/ftp/soa/>> [Consulta: julio 2015].
- Kobsa, A., 2001. Generic user modeling systems. *User Modeling and User-Adapted Interaction*, 11(1-2), pp. 49-63, 2001.
- Kobsa, A. and Fink, J., 2006. An LDAP-Based User Modeling Server and its Evaluation. *User Modeling and User-Adapted Interaction*, 16(2), pp. 129-169.
- Kuneva, M., 2009. Roundtable on Online Data Collection, Targeting and Profiling. Keynote Speech. [En línea] Disponible en: <http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm> [Consulta: noviembre 2015].
- Larman, C., 2001. *Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and the Unified Process* (2nd ed.). NJ: Prentice Hall PTR.
- Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, 2007. *Boletín Oficial Del Estado*, 276, pp. 47160-47165. Disponible en: <http://www.boe.es/diario_boe/txt.php?id=BOE-A-2007-19814> [Consulta: junio 2015]
- Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, 1999. *Boletín Oficial Del Estado*, 298. Disponible en: <<http://www.boe.es/buscar/act.php?id=BOE-A-1999-23750>> [Consulta: junio 2015].
- Lim, E. and Morgenstern, J., Buy.com. 2005. Method and system for Word of mouth advertising via a communications network. United States Patent Application 20050149397, Application number 10/997121.
- López, R., 2012. Innovación del modelo de negocio: propuesta de un modelo holístico. Tesis doctoral. Universidad Autónoma de Madrid. Disponible en: <<https://repositorio.uam.es/handle/10486/7825>> [Consulta: mayo 2015].
- Madden, M. and Rainie, L., 2015. Americans' Attitudes About Privacy, Security and Surveillance. Pew Research Center. Disponible en: <http://www.pewinternet.org/files/2015/05/Privacy-and-Security-Attitudes-5.19.15_FINAL.pdf> [Consulta: noviembre 2015].
- Madsen, P., Mont M. C. and Wilton, R., 2006. A Privacy Policy Framework. A position paper for the W3C Workshop of Privacy Policy Negotiation P3P. Ispra, Italy, 17-10 October. Hosted by the Joint Research Center (JRC) of the European Commission.
- Malekzadeh, B., 2010. Event-Driven Architecture and SOA in collaboration – A study of how Event-Driven Architecture (EDA) interacts and functions within Service-Oriented

- Architecture (SOA). Master of Thesis in IT Management. University of GothenBurg. Disponible en: https://gupea.ub.gu.se/bitstream/2077/23491/1/gupea_2077_23491_1.pdf [Consulta: octubre 2015].
- Martín García, Y.S., San Miguel González, B. and Yelmo García, J.C., 2009. Prosumers and accessibility: How to Ensure a Productive Interaction. Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibility (W4A). Madrid, España, 2009. ACM, pp. 50-53. [John M Slatin Award for Best Communication Paper].
- Martin, S. y Gil, L., 2012. Curso para operadores de Proveedores de Identidad 1.0. Confia. Disponible en: http://confia.aupa.info/docs/cursos/2012/idp/curso_administrador_idp.pdf [Consulta: julio 2015].
- Martínez C., 2009. Desarrollo de un Prototipo de Aplicación para Protección de Información Personal en Plataformas de Servicios Convergentes Centrados en el Usuario. Proyecto Fin de Carrera. Universidad Politécnica de Madrid.
- Michelson, B. M., 2006. Event-Driven Architecture Overview – Event-Driven SOA Is Just Part of the EDA Story. Boston: Patricia Seybold Group. Disponible en: <http://www.omg.org/soa/Uploaded%20Docs/EDA/bda2-2-06cc.pdf> [Consulta: septiembre 2015].
- Monjas, M.A., del Álamo, J.M., San Miguel, B. and Yelmo, J.C., 2010. Method for Privacy Management in an Identity Network, Physical Entities and Computer Programs Therefor. PCT/EP2009/054223, WO/2010/115460, US13263669 granted 2014. Disponible en: <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2010115460> [Consulta: junio 2015].
- Monjas, M.A., Del Álamo, J.M., San Miguel, B., Trapero, R. and Yelmo, J.C., 2012. Method for Selectively Distributing Information in a Computer or Communications Network, and Physical Entities Therefor. PCT/EP2010/059264, WO/2012/000543. EU National Phase: EP2589004. USA National Phase: US20130173712. Disponible en: <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2012000543> [Consulta: junio 2015].
- Montaner, M., López, B. and de la Rosa, J. L., 2003. A Taxonomy of Recommender Agents on the Internet. Artificial Intelligence Review, 19(4), pp. 285-330.
- Montero, J. M., 2014. Desarrollo de un prototipo de arquitectura EDA para el procesamiento y visualización de eventos de negocio en el ámbito bancario. Proyecto fin de carrera. Universidad Politécnica de Madrid.
- Naciones Unidas, 1948. Declaración Universal de los Derechos Humanos. Departamento de Información Pública de las Naciones Unidas. Disponible en: <http://www.un.org/es/documents/udhr/> [Consulta: septiembre 2015].

- Networked & Electronic Media (NEM), 2013. Big and Open data Position Paper. European Technology Platforms of Horizon 2020. Disponible en: <<http://nem-initiative.org/wp-content/uploads/2013/11/NEM-PP-016.pdf>> [Consulta: junio 2015].
- Niles, I. and Pease, A., 2001. Towards a Standard Upper Ontology. In: C. Welty and B. Smith, eds., 2001. 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), pp. 2-9.
- Nissenbaum, H., 2004. Privacy as contextual integrity. Washington Law Review, 79(1), pp. 119-157.
- Object Management Group (OMG), 2015. Financial Services Standards. [En línea] Disponible en: <<http://www.omg.org/hot-topics/finance.htm>> [Consulta: noviembre 2015].
- Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información (ONTSI), 2015. Estudio de caracterización del sector infomediario. Madrid: ONSI. Disponible en: <<http://www.ontsi.red.es/ontsi/es/informacion-y-recursos/estudio-de-caracterizaci%C3%B3n-del-sector-infomediario-2014>> [Consulta: mayo 2015].
- Open Data Institute y Fingleton Associates, 2014. Data Sharing and Open Data for Banks. A report for HM Treasury and Cabinet Office. Disponible en: <<https://www.gov.uk/government/publications/data-sharing-and-open-data-for-banks>> [Consulta: octubre 2015].
- OpenID Foundation, 2007. OpenID Authentication 2.0 – Final. Obsolete Specification.
- OpenID Foundation, 2014. OpenID Connect Core 1.0. Final Specification.
- Organisation for Economic Cooperation and Development (OECD), 2013. The OECD Privacy Framework. Disponible en: <http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf> [Consulta: julio 2015].
- Organisation for Economic Cooperation and Development (OECD), 2014. Protecting privacy in a data-driven economy: taking stock of current thinking. Working Party on Security and Privacy in the Digital Economy. Disponible en: <<http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/iccp/reg%282014%293&doclanguage=en>> [Consulta: junio 2015].
- Organization for the Advancement of Structured Information Standards (OASIS), 2005. Assertions and protocols for the OASIS Security Assertion Markup Language (SAML) Version 2.0. OASIS Standard. Disponible en: <<http://docs.oasis-open.org/security/saml/v2.0/>> [Consulta: julio 2015].
- Organization for the Advancement of Structured Information Standards (OASIS), 2005b. eXtensible Access Control Markup Language (XACML) Version 2.0. OASIS Standard. Disponible en: <http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf> [Consulta: julio 2015].

- Organization for the Advancement of Structured Information Standards (OASIS), 2009. Web Services Federation Language (WS-Federation) Version 1.2. OASIS Standard. Disponible en: <<http://docs.oasis-open.org/security/saml/v2.0/http://docs.oasis-open.org/wsfed/federation/v1.2/os/ws-federation-1.2-spec-os.html>> [Consulta: julio 2015].
- Osterwalder, A. and Pigneur, Y., 2010. Business Model Generation: a Handbook for Visionaries, Game Changers, and Challengers. New Jersey: John Wiley & Sons.
- Privacy by Design, 2011. Privacy by Design Curriculum 2.0. [En línea] Disponible en: <<https://www.privacybydesign.ca/index.php/publications/curriculum/>> [Consulta: septiembre 2015].
- Rappa, M., 2003. Business Models on the Web. Managing the Digital Enterprise. [En línea] Disponible en: <<http://digitalenterprise.org/models/models.html>> [Consulta: mayo 2015].
- Real Academia de Ingeniería, 2015. Definición de red social. [En línea] Disponible en: <<http://diccionario.raing.es/es/lema/red-social>> [Consulta: junio 2015].
- Ros-Martín, M., 2009. Evolución de los servicios de Redes Sociales en Internet. El Profesional de la Información, 18 (5), pp. 1019-1060.
- Ross, W., 2014. How Much Is Your Privacy Worth?. MIT Technology Review. [En línea] Disponible en: <<http://www.technologyreview.com/news/529686/how-much-is-your-privacy-worth/>> [Consulta: junio 2015].
- San Miguel, B., Aguirre, A., del Álamo, J. M. and Cortes, M., 2012. A Proposal for Enhancing the Motivation in Students of Computer Programming. 5th International Conference of Education, Research and Innovation (ICERI 2012). Madrid, Spain. 19-21 Nov 2012.
- San Miguel, B., del Alamo, J.M. and Yelmo, J.C., 2014a. Personal Data Usage and Management. Doctoral Consortium on Enterprise Information Systems (DCEIS 2014). Lisbon, Portugal. 27-30 April 2014. Scitepress, pp. 26-29.
- San Miguel, B., del Alamo, J.M. and Yelmo, J.C., 2014b. Evolving a Core Banking Enterprise Architecture: Leveraging Business Events Exploitation. 16th International Conference on Enterprise Information Systems (ICEIS 2014). Lisbon, Portugal, 27-30 April 2014. Scitepress, pp. 181-189. [Best paper award]
- San Miguel, B., Del Álamo, J.M. and Yelmo, J.C., 2014c. Creating and Modelling Personal Socio-Economic Networks in On-Line Banking. In: B. Benatallah, et al., eds. 2014. Web Information Systems Engineering – WISE 2014 Workshops: 15th International Workshops IWCSN 2014, Org2 2014, PCS 2014, and QUAT 2014, Thessaloniki, Greece, October 12-14, 2014, Revised Selected Papers. Springer International Publishing. Lecture Notes in Computer Science, 9051, pp. 177-190. [Revised Selected Papers]
- San Miguel, B., Del Álamo, J.M. and Yelmo, J.C., 2015a. A Personal Data Framework for Exchanging Knowledge about Users in New Financial Services. 1st International Workshop on Personalization & Recommender Systems in Financial Services (FINREC). Graz, Austria, 16 April 2015. CEUR, 1349, pp. 19-26.

- San Miguel, B., Cases, N. N., Yelmo, J.C. and Del Álamo, J.M., 2015b. Approaching Unbanked People: A Person-to-Person Payment Application on Facebook. VII Congreso Iberoamericano de Telemática (CITA2015). Popayán, Colombia. 10 June 2015.
- San Miguel, B., Del Álamo, J.M. and Yelmo, J.C., 2015c. Extracting Knowledge through Friendsourcing: Application to new Financial Services. VII Congreso Iberoamericano de Telemática (CITA2015). Popayán, Colombia. 10 June 2015. [Selected Best Papers]
- San Miguel, B., Del Álamo, J.M. and Yelmo, J.C., 2015d. Integrating Business Information Streams in a Core Banking Architecture: A Practical Experience. In: J. Cordeiro, et al., eds. 2015. Enterprise Information Systems: 16th International Conference, ICEIS 2014, Lisbon, Portugal, April 27-30, 2014, Revised Selected Papers. Springer International Publishing. Lecture Notes in Business Information Processing, 227, pp. 417-433.
- San Miguel, B., Del Álamo, J.M. and Yelmo, J.C., 2015e. A Personal Data Framework for Distributed User Modelling in New Service Ecosystems. IEEE América Latina, 13(11), pp. 3604-3611.
- Simcott, R., 2014. Social Media Fasts: Latin America. Blog Emoderation Social Media Management, April 2, 2014. [En línea] Disponible en: <<http://www.emoderation.com/social-media-fast-facts-latin-america/>> [Consulta: octubre 2015].
- Solove, D. J., 2006. A Taxonomy of Privacy. University of Pennsylvania Law Review, 154(3), pp. 477-560.
- Sommerville, I., 2006. Software Engineering (9th ed.). Boston: Addison-Wesley.
- Sosnovsky, S. and Dicheva, D., 2010. Ontological Technologies for User Modelling. International Journal of Metadata, Semantics and Ontologies, 5(1), pp. 32-71.
- Spiekermann, S., 2009. Engineering Privacy. IEEE Transactions on Software Engineering, 35(1), pp. 67-82.
- Su, X. and Khoshgoftaar, T. M., 2009. A Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence, 2009, article 4.
- Tang, H. and Song, L., 2011. Ontologies in financial services: Design and applications. International Conference on Business Management and Electronic Information. Guangzhou, China, 13-15 May 2011. IEEE, pp. 364-367.
- Taylor, H., Yochem, A., Phillips, L. and Martinez, F., 2009. Event-Driven Architecture: How SOA Enables the Real Time Enterprise (1st ed.). Boston: Addison-Wesley Professional.
- The Statistics Portal, 2015. Number of Internet Users in Latin America from 2013 to 2018. [En línea] Disponible en: <<http://www.statista.com/statistics/274860/number-of-internet-users-in-latin-america/>> [Consulta: octubre 2015].
- Timmers, P., 1998. Business Models for Electronic Markets. Electronic Markets, 8 (2), pp. 3-8.

- Toca, G., 2014. Data Brokers: El gran negocio de la información privada. Forbes, September 2014, pp. 166-171.
- Twitter, 2015. Twitter Anuncios. Anúnciese en Twitter. [En línea] Disponible en: <https://ads.twitter.com/login?redirect_after_login=https%3A%2F%2Fads.twitter.com%2F> [Consulta: junio 2015].
- Uschold, M. and Gruninger, M. 1996. Ontologies: Principles, methods and applications. Knowledge Engineering Review, 11(2), pp. 93-136.
- Van der Sluijs, K. and Houben, G. J., 2005. Towards a Generic User Model Component. Workshop on Personalisation on the Semantic Web held in conjunction with the 10th International Conference on User Modeling (UM2005). Edinburgh, Scotland, pp. 47-57.
- Viviani, M., Bennani, N. and Egyed-Zsigmond, E., 2010. A Survey on User Modeling in Multi-application Environments. Proceedings of the 2010 Third International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services, pp. 111-116.
- Weinberg, T., 2009. The New Community Rules: Marketing On The Social Web. Estados Unidos: O'Reilly.
- Westin, A. F., 1968. Privacy and Freedom. Washington and Lee Law Review, 25(1), pp. 166-170.
- Wikipedia, 2015. Criticism of Google. [En línea] Disponible en: <http://en.wikipedia.org/wiki/Criticism_of_Google> [Consulta: junio 2015]
- World Economic Forum (WEF), 2011. Personal Data: The Emergence of a New Asset Class. Disponible en: <<http://www.weforum.org/reports/personal-data-emergence-new-asset-class>> [Consulta: mayo 2015].
- World Wide Web (W3C), 2004. OWL Web Ontology Language. W3C Recommendation. Disponible en: <<http://www.w3.org/TR/2004/REC-owl-features-20040210/>> [Consulta: septiembre 2015].
- World Wide Web (W3C), 2012. The WebSocket API. Candidate Recommendation. Disponible en: <<http://www.w3.org/TR/2012/CR-websockets-20120920/>> [Consulta: noviembre 2015].
- World Wide Web (W3C), 2014. OpenSocial Foundation Moves Standards Work to W3C Social Web Activity. [En línea] Disponible en: <<http://www.w3.org/blog/2014/12/opensocialfoundation-moves-standards-work-to-w3c-social-web-activity/>> [Consulta: junio 2015].
- World Wide Web Consortium (W3C), 2015. SPARQL Current Status. [En línea] Disponible en: <http://www.w3.org/blog/2014/12/opensocialfoundation-moves-standards-work-to-w3c-social-web-activity/http://www.w3.org/standards/techs/sparql#w3c_all> [Consulta: septiembre 2015].

- Yelmo García, J.C., Martín García, Y.S., San Miguel González, B. y Albors Fernández I., 2006. Herramienta de evaluación de accesibilidad de contenidos y aplicaciones para terminales móviles. XVI Jornadas Telecom I+D. Madrid.
- Yelmo García, J.C., Martín García, Y.S. y San Miguel González, B., 2007a. Contribución a las metodologías de evaluación de accesibilidad de contenidos para la Web Móvil. CEDI 2007 – II Congreso Español de Informática (MWeb '07 Jornadas Sobre Web Móvil). Zaragoza, 2007.
- Yelmo García, J.C., San Miguel González, B. y Martín García, Y.S., 2007b. Análisis del potencial de las comunicaciones móviles para la mejora de los servicios proporcionados por la Administración Pública a las personas con discapacidad. Tecnimap 2007 – X Jornadas sobre Tecnologías de la Información para la Modernización de las Administraciones Públicas. Gijón, 2007.
- Yelmo, J. C., Trapero, R. and del Alamo, J. M., 2009. Identity Management and Web Services as Service Ecosystem Drivers in Converged Networks. IEEE Communications Magazine, 47(3), pp.174-180.
- Yelmo García, J.C., Martín García, Y.S. and San Miguel González, B., 2010. Semantic Characterization of Context of Use and Contents for User-Centric Media Retrieval. In: F. Álvarez and C. Costa, eds. 2010. User Centric Media: Second International ICST Conference, UCMedia 2010, Palma de Mallorca, Spain, September 1-3, 2010. Revised Selected Papers. Springer Berlin Heidelberg. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 60, pp 20-25.
- Yelmo García, J.C., Martín García, Y.S. and San Miguel González, B., 2011a. Modelado Semántico y Centrado en el Usuario de Servicios Adaptados al Contexto de Uso. VI Congreso Ibero-americano de Telemática (CITA2011). Gramados, Brasil. 16-18 mayo 2011. Cadernos de Informática, 6(1).
- Yelmo García, J.C., San Miguel González, B. y Martín García, Y.S., 2011b. El Rol Central del Usuario en los Servicios Telemáticos. Buscamedia White papers. [En línea] Disponible en: <http://www.cenitbuscamedia.es/index.php?option=com_rokdownloads&view=file&Itemid=59&id=43> [Consulta: diciembre 2015].
- Ying, W., Sujanani, A., Ray, P., Paramesh, N., Lee D. and Bhar, R., 2009. Design and development of financial applications using ontology-based multi-agent systems. Computing and Informatics, 28(5).
- Yu, L., 2011. A Developer's Guide to the Semantic Web. Heidelberg: Springer.
- Zhang, Y. and Yu, T., 2012. Mining Trust Relationships from Online Social Networks. Journal of Computer Science and Technology, 27(3), pp.492–505.
- Zott, M., Amit, R. H. and Massa, L., 2011. The Business Model: Theoretical Roots, Recent Developments, and Future Research. Journal of Management, 37(4), pp. 1019-1060.