UNIVERSIDAD DE SEVILLA

Escuela Superior de Ingenieros

Aportaciones Metodológicas a la Integración de Sistemas de Información Sanitarios basada en Gestión Semántica

TESIS DOCTORAL

Isabel Román Martínez

Ingeniero de Telecomunicación

Sevilla, 2006

Departamento de Ingeniería de Sistemas y Automática

Área de Telemática

Escuela Superior de Ingenieros

UNIVERSIDAD DE SEVILLA

Aportaciones Metodológicas a la Integración de Sistemas de Información Sanitarios basada en Gestión Semántica

TESIS DOCTORAL

Autor:

Isabel Román Martínez

Ingeniero de Telecomunicación

Directores:

Laura María Roa Romero

Profesora Doctora en Ingeniería de Sistemas y Automática

Germán Madinabeitia Luque

Doctor Ingeniero de Telecomunicación

Sevilla, 2006

Tribunal nombrado por el Magnífico y Excelentísimo Sr. Rector de la Universidad de Sevilla, el día de 2006.

Presidente:

Vocales:

Secretario:

Suplente:

Realizado el acto de defensa y lectura de la Tesis el día

En Sevilla, acuerda otorgarle la calificación de

El Presidente **Los Vocales**

El Secretario

Agradecimientos

En primer lugar quiero dar las gracias a mis directores de Tesis por su dirección y guía durante estos años y su disponibilidad perenne para ayudarme y apoyarme de cualquier modo posible. Laura María Roa Romero ha sido el alma y guía no sólo de esta Tesis Doctoral sino de todas las actividades de investigación que he llevado a cabo y su apoyo continuo me ha demostrado que lo seguirá siendo por muchos años. Germán Madinabeitia Luque ante todo me ha enseñado a valorar el trabajo bien hecho, aunque el esfuerzo para conseguirlo no se vea recompensado a corto plazo. Los dos han tenido la virtud de combinar el apoyo científico con la amistad desinteresada y de hacerme mejorar como persona día a día.

En segundo lugar a todos los compañeros del Área de Ingeniería Telemática que me han ayudado en mis tareas docentes e investigadoras durante estos años ofreciéndome al mismo tiempo su valiosa amistad. En especial a Juan Manuel Vozmediano, que confió en mí para formar parte de este maravilloso equipo, a Juan Antonio Ternero y Teresa Ariza, que siempre están disponibles cuando se les necesita y a Jose Ángel Gómez, que se desvive por todos desinteresadamente.

En tercer lugar a los miembros y colaboradores del Grupo de Ingeniería Biomédica. En especial a los Doctores Alfonso Palma y Antonio Milán, por todo el tiempo dedicado fuera de sus obligaciones asistenciales a la investigación en Ingeniería Biomédica, y a Javier Reina y Manuel Prado por su inestimable ayuda en incontables ocasiones y por su amistad y apoyo continuos.

Mi agradecimiento también a todos los investigadores que han participado en el proyecto MEDICI y que me han ayudado en distintas tareas de mi investigación. En especial a Manuel Campos, José Palma y Roque Martín por su calurosa acogida y su apoyo durante mi estancia en Murcia.

También quiero agradecer su interés por mi trabajo, su apoyo y su disponibilidad de ayuda a miembros de otros grupos de investigación. A Maria Elena Hernando, Enrique Gómez, César Cáceres, Ángel García-Olaya y el resto de miembros del GBT en la ETSIT de la Universidad Politécnica de Madrid, porque siempre me han tratado como una amiga y me han dado valiosos consejos. A Montserrat Robles, Pere Crespo y Jose

Alberto Maldonado del BET de la Universidad Politécnica de Valencia, que me han animado siempre a continuar con mi trabajo. A Asunción Gómez del Departamento de Inteligencia Artificial de la Facultad de Informática en la Universidad Politécnica de Madrid, por sus recomendaciones durante la escuela de verano de Web Semántica. A Thomas Beale y los miembros de la comunidad del openEHR por el interés mostrado por mi trabajo y las prontas respuestas a las dudas planteadas en la lista de distribución. Y finalmente a Ana Belén García, Jorge López de Vergara y el resto de miembros del DIT en la ETSIT de la Universidad Politécnica de Madrid, por su apoyo, su amistad y su siempre calurosa acogida.

Para terminar agradecer el apoyo de toda mi familia y amigos. A mis padres Carmen y Francisco que han dedicado toda su vida a sus hijos. A mi hermano Isco que se lo merece todo. A Susana que no me ha fallado jamás. Y un agradecimiento muy especial a Antonio que es la luz de mi camino.

A mi madre

Índice de Contenidos

Agradecimientos *7*

Índice de Contenidos *11*

Índice de Figuras *16*

Índice de Tablas *19*

Glosario *21*

Resumen *25*

Summary *27*

Capítulo 1. Introducción *29*

1.1. Introducción **29**

1.1.1. Sistemas de Información Sanitarios 29

1.1.2. La Historia Clínica 30

1.1.3. La Historia Clínica Electrónica 31

1.1.4. La gestión de la historia clínica 33

1.1.5. Tecnología en la historia clínica 33

1.1.6. La historia clínica en el conjunto de la actividad sanitaria 34

1.2. Motivación **36**

1.2.1. Arquitectura objetivo 36

1.2.2. El problema de la federación de bases de conocimiento 37

1.3. Objetivos **39**

1.4. Estructura de la memoria **40**

Capítulo 2. Normalización *43*

2.1. Introducción **43**

2.2. Estrategias de normalización para el interfuncionamiento **45**

2.3. Organismos oficiales de normalización **46**

2.3.1. CEN/TC251 “Health Informatics” 46

2.3.2. ISO/TC215 "Health Informatics" 48

2.3.3. Health Level Seven: HL7 49

2.4. Otros organismos de normalización **50**

2.4.1. ACR-NEMA (DICOM) 50
2.4.2. The Institute of Electrical and Electronics Engineers (IEEE) 53
2.4.3. CORBAmed del OMG 54
2.4.4. Estándares de terminología 55

2.5. Iniciativas de apoyo a estándares 56
2.5.1. OpenEHR 56
2.5.2. Integrating the Healthcare Enterprise: IHE 56

2.6. Conclusión 57

Capítulo 3. Técnicas de semántica 59

3.1. Introducción 59

3.2. Plataformas de integración 60
3.2.1. Objetos distribuidos 60
3.2.2. Agentes Software 61
3.2.3. Servicios Web 62
3.2.4. Grid 64

3.3. Descripción semántica del dominio: Ontologías 64

3.4. Integración basada en la gestión semántica 66
3.4.1. La Web Semántica 66
3.4.2. Aportaciones y limitaciones de la Web Semántica a la integración de sistemas de información sanitarios. 69
3.4.3. La Grid Semántica 71
3.4.4. Aportaciones y limitaciones de la Grid Semántica a la integración de sistemas de información sanitarios. 73

3.5. Conclusiones 73

Capítulo 4. Métodos y modelos 75

4.1. Introducción 75

4.2. El modelo dual 75
4.2.1. Introducción 75
4.2.2. Arquetipos 79
4.2.3. Conclusiones 82

4.3. Open Distributed Processing 83
4.3.1. Introducción 83
4.3.2. El modelo de referencia ODP (RM-ODP) 86
4.3.3. Vistas 88
4.3.4. Algunos conceptos importantes en ODP 90

4.4. Una especialización de ODP: la norma HISA....97
4.4.1. Introducción....97
4.4.2. Vista de empresa....99
4.4.3. Vista de información....101
4.4.4. Vista computacional....101
4.4.5. Conformidad....102
4.4.6. Conclusiones....104

Capítulo 5. Servicio demográfico....107

5.1. Introducción....107
5.1.1. Los servicios demográficos....107
5.1.2. Identificación de personas....108
5.1.3. Recuperación de información demográfica....109

5.2. Modelos de datos....109
5.2.1. OpenEHR....109
5.2.2. CEN....114
5.2.3. PID....117

5.3. Servicio de identificación de personas en CORBAmed....120
5.3.1. Conceptos básicos....120
5.3.2. Interfaces....121

5.4. Conclusiones....126

Capítulo 6. Resultados: extensiones a HISA....129

6.1. Introducción....129

6.2. Paradigma Estratégico....130
6.2.1. Requisitos generales....130
6.2.2. Consideraciones generales de arquitectura. Extensión de los términos HISA.....134
6.2.3. Consideraciones de integración....135
6.2.4. Consideraciones sobre el papel de las aplicaciones y componentes de gestión de historia clínica dentro de la arquitectura HISA....136

6.3. Extensión de la vista de empresa....138
6.3.1. Reflejo del modelo federado en la vista de empresa....142
6.3.2. Flujo de trabajo orientado al sujeto de cuidado (paciente)....143
6.3.3. Flujo de información clínica....146
6.3.4. Flujo de gestión de actividad....149
6.3.5. Gestión de recursos de la organización sanitaria....153
6.3.6. Actividades de seguridad y gestión de recursos....154
6.3.7. Gestión de semántica....160

6.3.8. Gestión de comunicaciones....164
6.3.9. Consideraciones de los sistemas de historia clínica única en la vista de empresa....166
6.3.10. Consideraciones finales y conclusiones....168

6.4. Extensión de la vista de información....169
6.4.1. Introducción....169
6.4.2. Modificación del modelo de semántica....178
6.4.3. Extensión del modelo del sujeto de cuidado....187
6.4.4. Extensión del modelo de información clínica....190
6.4.5. El modelo de gestión de actividad....194
6.4.6. Modelo de recursos de organización....194
6.4.7. Modelos de Seguridad y Gestión de recursos....195
6.4.8. Modelo de comunicaciones....201
6.4.9. Conclusiones....202

6.5. Extensión de la vista computacional....202
6.5.1. Introducción....202
6.5.2. Lenguaje computacional....206
6.5.3. Componentes semánticos de la arquitectura base....216
6.5.4. Componentes de gestión semántica....218
6.5.5. Componentes relacionados con el sujeto de cuidado....220
6.5.6. Componentes de información clínica....222
6.5.7. Componentes de gestión de actividad....225
6.5.8. Componentes de recursos de organización....226
6.5.9. Componentes de seguridad y gestión de recursos....227
6.5.10. Componentes de comunicaciones....230
6.5.11. Conclusiones....233

6.6. Aproximaciones a las vistas de ingeniería y tecnológica....233

Capítulo 7. Validación: gestión demográfica....235

7.1. Paradigma estratégico....235

7.2. Vista de empresa....245
7.2.1. Introducción....245
7.2.2. Actividades de gestión de personas sobre sistemas autónomos....247
7.2.3. Actividades de gestión de personas sobre la federación....250

7.3. Vista de información....253
7.3.1. El modelo de referencia o la ontología base....257
7.3.2. La ontología del dominio....261

7.4. Vista computacional....264
7.4.1. Interfaces básicas y autónomas....265

7.4.2. Interfaces complejas y autónomas 274
7.4.3. Interfaces federadas 275
7.4.4. Comportamiento 278
7.4.5. Algunas consideraciones finales 285

Capítulo 8. Contribuciones originales 287

Capítulo 9. Publicaciones originales 293

9.1. Año 2006 293
9.1.1. Publicaciones en revistas 293

9.2. Año 2005 293
9.2.1. Capítulos de libro 293
9.2.2. Actas de congresos 293

9.3. Año 2004 294
9.3.1. Actas de congresos 294

9.4. Año 2003 295
9.4.1. Actas de congresos 295

9.5. Año 2002 296
9.5.1. Actas de congresos 296

9.6. Año 2001 296
9.6.1. Actas de congresos 296

Capítulo 10. Línea de investigación 297

Índice de Figuras

Figura 1. Modelo de comunicaciones en DICOM (extraído de [51]) 51

Figura 2. Algunos servicios definidos por CORBAmed 55

Figura 3. Modelo de servicio en OWL-S 68

Figura 4. Arquitectura de la GRID Semántica 72

Figura 5. Arquitectura del modelo de arquetipos 82

Figura 6. Tipos de interacciones en ODP 92

Figura 7. Vinculación primitiva 93

Figura 8. Vinculación compuesta 94

Figura 9. Objeto compuesto a partir de la interacción de otros 95

Figura 10. Relación entre servicios básicos, compuestos y aplicaciones 96

Figura 11. Papel de HISA en la arquitectura sanitaria 98

Figura 12. Modelo de información demográfico del OpenEHR (extraída de [33]) 110

Figura 13. Modelo de información para identificadores en el openEHR (extraída de [66]) 111

Figura 14. Tipo de dato para acomodar identificadores en el OpenEHR. 113

Figura 15. Tipos de Datos básicos dados por el CEN/TC251 114

Figura 16. Resumen de información demográfica en los GPICS no clínicos 115

Figura 17. Modelo de información para identificadores en el PID de CORBAmed. 118

Figura 18. Modelo de información demográfica en el PID de CORBAmed 119

Figura 19. Interfaces del Servicio PID 122

Figura 20. Modificaciones propuestas a la clasificación de actividades en HISA 141

Figura 21. Principales componentes del flujo de trabajo del sujeto de cuidado. 146

Figura 22. Principales componentes del flujo de trabajo de gestión de información clínica. 148

Figura 23. Principales componentes del flujo de gestión de actividad. 153

Figura 24. Principales componentes de seguridad en la arquitectura sanitaria 155

Figura 25. Principales componentes de gestión de semántica..... 161

Figura 26. Representación de algunas actividades de gestión de HCE y su relación con actividades presentadas en la vista de empresa. 166

Figura 27. Relación entre objetos descriptivos y operacionales..... 174

Figura 28. Estructura general de las clases en la vista de información de HISA 176

Figura 29. Modelo de conceptos genéricos en ontologías..... 182

Figura 30. Modelo de reglas..... 184

Figura 31. Modelo simplificado para anotaciones..... 186

Figura 32. El objeto persona en un modelo federado..... 189

Figura 33. Relación entre los modelos de información del COAS y del EN13606 (Simplificado)..... 191

Figura 34. Modelo para la Gestión de usuarios, identificación y credenciales 197

Figura 35. Modelo de información de políticas de acceso del EN13606-4..... 199

Figura 36. Modelo de información del RAD de CORBAmed 200

Figura 37. Clases objeto computacional e interfaz..... 208

Figura 38. Clase servicio 208

Figura 39. Subtipos de vinculación 209

Figura 40. Clases de firma y datos 210

Figura 41. Tipos de interfaz para la vinculación 211

Figura 42. Datos para la clasificación de métodos básicos 213

Figura 43. Datos para la clasificación de métodos complejos..... 214

Figura 44. Datos para la clasificación de actividades..... 215

Figura 45. Estados posibles de un contacto..... 221

Figura 46. Vistas de la arquitectura..... 237

Figura 47. Centralización de la gestión demográfica 240

Figura 48. Gestión semi-centralizada de la información demográfica..... 241

Figura 49. Federación de la gestión demográfica.....243

Figura 50. Modelo colaborativo para la gestión de información demográfica.....244

Figura 51. Vista de información. Ontología de referencia y del dominio.....256

Figura 52. Ejemplo de arquetipo de TraitSpec.....264

Figura 53. Ejemplo de interfaces con causalidad complementaria. 270

Figura 54. Otras interfaces de ejemplo, todas con causalidad servidor.....270

Figura 55. Definición de servicios de gestión de entidad y de gestión de atributos.....271

Figura 56. Descripción del Servicio para gestión de entidades.....272

Figura 57. Resumen de la especificación de la operación para listar entidades (pertenece a la interfaz del servicio de gestión de entidades).....273

Figura 58. Arquitectura diseñada para la gestión federada de identificadores y datos demográficos 280

Figura 59. Procedimiento para la obtención del perfil genérico federado.....282

Figura 60. Procedimiento para la identificación de la persona en la federación.....283

Figura 61. Asentamiento de candidato y obtención de perfil o rol de entidad.....285

Índice de Tablas

Tabla I. . Clasificación de interfaces 265

Tabla II. Parámetros definidos para la gestión demográfica 266

Tabla III. Terminaciones 267

Tabla IV. Operaciones (tipo interrogación) e invocaciones 268

Glosario

ACL: Agent Communication Language

ACR: American Collage of Radiology

ADL: Archetype Definition Language

ADSL: Asymmetric Digital Subscriber Line

ANSI: American National Standards Institute

CDA: Clinical Document Architecture

CEN: Comité Europeo de Normalización

CENELEC: Comité Europeo de Normalización Electrotécnica

CIAS: Clinical Image Access Service

COAS: Clinical Observation Access Service

CORBA: Common Object Reques Broker Architecture

CORBAmed: Common Object Reques Broker Architecture Medical

DAI: Distrituted Artificial Intelligence

DICOM: Digital Imaging and COmmunications in Medicine

DNI: Documento Nacional de Identidad

EHR: Electronic Healthcare Record

EN: European Standard

ENV: European Prestandard

EUGLOSS: Multilingual glossary of technical and popular medical terms in nine European languages

FIPA: Foundation for Intelligent Physical Agents

GEHR: Good European Health Record

GPIC: General Purpose Components for messages

HC: Historia Clínica

HCE: Historia Clínica Electrónica

HIS: Hospital Information System

HISA: Healthcare Information System Architecture

HL7: Health Level Seven

ICD: International Classification of Diseases

ID: Identificador

IDL: Interface Description Language

IEEE: The Institute of Electrical and Electronics Engineers

IEEE-SA: Institute of Electrical and Engineers Standards Association

IHE: Integrating the Healthcare Enterprise

IP: Identificador de Persona

ISO: International Organization for Standardization

ITU-T: International Telecommunication Union. Telecommunication Standardization Sector

IUPAC: International Union of Pure Applied Chemistry

KIF: Knowledge Interchange Format

KQML: Knowledge Query and Manipulation Language

KSE: Knowledge Sharing Effort

LOINC: Laboratory Observation Identifier, Names and Codes

LQS: Lexicon Query Service

NEMA: National Electrical Manufactureres Association

ODP: Open Distributed Processing

OMG: Object Management Group

OpenEHR: Open Electronic Healthcare Record

OSF: Open Software Foundation

OWL: Ontology Web Language

OWL-S: Web Ontology Language for Services

PACS: Picture Archiving and Communications Systems

PIDS: Person Identification Service

PMI: Person Master Index

QoS: Quality of Service

RAD: Resource Access Decision

RDF: Resource Description Framework

RIM: Reference Information Model

RIS: Radiology Information System

RM-ODP: Reference Model- Open Distributed Processing

RTC: Red Telefónica Conmutada

SAML: Security Assertion Markup Language

SDO: Standards Development Organization

SNOMED: Systematized Nomenclature of Medicine

SOAP: Simple Object Access Protocol

SWRL: Semantic Web Rule Language

TCP/IP: Transmission Control Protocol/Internet Protocol

TIC: Tecnologías de la Información y las Comunicaciones

TINA: Telecommunication Information Networking Architecture

UDDI: Universal Description, Discovery and Integration

UML: Unified Modeling Language

UMLS: Unified Medical Language System

URI: Uniform Resource Identifiers

URL: Uniform Resource Locator

WG: Working Group

WSDL: Web Services Description Language

WSDL-S: Web Services Description Language - Semantic

WSML: Web Services Modeling Language

XML: eXtensible Markup Language.

Resumen

La aplicación de las tecnologías de la información y las comunicaciones está provocando un cambio significativo en la ejecución de las actividades llevadas a cabo en el seno de las organizaciones sanitarias. Esta implantación tecnológica se ha venido realizando a medida que se desarrollan nuevos avances y para cubrir necesidades específicas sin considerar, por lo general, la elaboración de una estrategia más compleja en la que se planifique la integración de los diversos sistemas que se despliegan. Esto provoca que, en la mayoría de las ocasiones, dentro de una organización sanitaria se tengan numerosos sistemas aislados que ofrecen unas prestaciones específicas pero que gestionan la información de forma completamente autónoma.

Dado que el principal servicio ofrecido por estas organizaciones es el asistencial, es precisamente en el desempeño de estas tareas donde primero surge la necesidad de integración. Una visión centrada en la atención conlleva el enfoque tradicional centrado en el paciente, que presenta como prioridad el desarrollo de la historia clínica electrónica única para permitir la gestión integrada de la información de salud relativa al paciente y distribuida en distintos sistemas. Debido a esto los mayores esfuerzos en el desarrollo de métodos, estrategias y normas de integración de sistemas de información sanitarios se están realizando precisamente en el ámbito de la historia clínica electrónica.

Sin embargo, las actividades en una organización sanitaria van más allá de las asistenciales, aunque éstas sean las entendidas como servicio final ofrecido al cliente. En la mayoría de las ocasiones se puede encontrar algún tipo de vínculo entre estas actividades que puede facilitar la explotación de las sinergias existentes entre las mismas. Esto da lugar a una nueva visión del problema de integración de sistemas que está centrada en la gestión eficiente del conocimiento de la organización.

En esta Tesis Doctoral se presenta una metodología para la integración de sistemas sanitarios que sigue este nuevo paradigma, en el que la actividad asistencial se entiende como una más dentro de la organización y ésta, como el resto, puede verse enriquecida si se trata adecuadamente el conocimiento almacenado en el conjunto de sistemas.

La Tesis está enmarcada en el contexto europeo, por lo que las aportaciones metodológicas se integran en la prenorma prEN 12967 del CEN. De manera que los

resultados se presentan como contribuciones a las vistas de empresa, de información y computacional, conforme al modelo ODP. Éstas se pueden resumir como:

- Desarrollo del paradigma estratégico para una arquitectura genérica de sistemas de información sanitarios. Destacando el papel de los sistemas de HCE y el paradigma de federación.
- Aportaciones a la vista de empresa, con atención especial a los requisitos necesarios para soportar la gestión semántica.
- Modificación del paradigma de modelado de la vista de información, atendiendo al modelo dual, y aportaciones a los modelos de información de primer nivel.
- Contribuciones a la vista computacional, donde se destaca la propuesta de un lenguaje computacional abierto, basado en ODP, para la especificación de componentes.

Como escenario de validación, la metodología propuesta se aplicará a las actividades relativas a la gestión demográfica y en especial a las de identificación de personas en una federación de sistemas, primer punto crítico en la integración de sistemas de HCE. En este sentido se ha desarrollado el siguiente trabajo:

- Descripción del paradigma estratégico, considerando especialmente el modelo de federación de sistemas.
- Desarrollo de la vista de empresa, describiendo los vínculos entre las tareas de identificación y gestión de datos personales y el resto de tareas de la organización.
- Especificación de la vista de información como modelos ontológicos de referencia y del dominio.
- Especificación de la vista computacional utilizando el lenguaje computacional desarrollado.

Summary

Many healthcare organizations are increasing their attention on using information and communication technologies (ICT). Services and work practices to which ICT are directly involved could be significantly improved.

Most healthcare providers, such as hospitals, have already been computerized at isolated departmental level, when particular needs appear and usually without the consideration of a global strategy for integration. Under the present circumstances, the large number of applications mutually isolated and incompatible, that are already available in the market, installed and operational in healthcare organizations, effectively supporting specific needs of users, cannot be underestimated.

Integration needs show up first in the patient care activities. The traditional patient centred view makes that improving access to clinical records is identified as a strategic aim. The development of electronic healthcare records (EHR) could facilitate the integrated management of the most relevant clinical information, accessing and retrieving it from different systems, whenever and wherever needed. Is just in the EHR field where the main efforts in the development of methodologies, strategies and standards are being done. But many other activities are executed in a healthcare organization. Some links between these activities could be frequently identified. The correct management of these links could improve the exploitation of synergies inside the organization. This analysis of the integration problem supposes a new paradigm, centred in an optimum management of the organization knowledge.

This PhD work introduces a methodology for healthcare systems integration following this new paradigm, where care activities are considered as any other inside the organization, and could be enriched with the correct management of the knowledge stored in the set of systems.

Any contribution to healthcare integration issues inside the European context should be compliant with prEN12967 norm and, consequently, with ODP methodology. The current methodological results are shown as contributions to enterprise, information and computational viewpoints from prEN12967. These contributions can be summarized as:

- Development of the strategic paradigm for a generic healthcare information systems architecture. The role of EHR and the federation approach are specially considered.

- Contributions to enterprise viewpoint, with attention to semantic management requirements.

- The information viewpoint has been revised including the dual model approach and designing the first level information model.

- In the computational viewpoint, an open computational language, based in ODP, is proposed for component's specification.

Modelling of activities related to demographic management has been used for the proposed methodology validation. Person identification in a federated scenario has been specially considered. These activities repeatedly arise as a priority for achieving integration. The following validation tasks have been done:

- Description of the strategic paradigm for demographic and identification activities, considering federation approach.

- Development of the enterprise viewpoint, describing links between identification and demographic management and the other activities inside the organization.

- Specification of the information viewpoint as reference and domain ontological models.

- Specification of the computational viewpoint using the developed computational language.

Capítulo 1. Introducción

1.1. Introducción

1.1.1. Sistemas de Información Sanitarios

El uso de las Tecnologías de la Información y las Comunicaciones (TIC) en el entorno sanitario mejora considerablemente la eficiencia de la gestión de la información de interés en una organización sanitaria [1-3]. Podría definirse un sistema de información sanitario como el conjunto de medios que permiten el almacenamiento, extracción y evaluación de la información útil en el entorno sanitario.

Los sistemas de información orientados al paciente son aquellos en los que se maneja información relativa a un individuo en particular con fines de gestión, diagnosis o terapia. Entre éstos se destacan los sistemas de radiología e imágenes médicas, de laboratorio o los de gestión de historia clínica [4]. Aunque la implantación de éstos mejora claramente la actividad asistencial, no son los únicos de interés en el contexto sanitario, donde también se deben considerar otras tareas como la investigadora, formativa, financiera o administrativa. Entre la información manejada en todos estos subsistemas existen vínculos que hacen prever que el uso de una estrategia global de implantación de las TIC en el seno de una organización sanitaria permitiría la optimización de esfuerzos y resultados.

El modelo administrativo actual, altamente descentralizado, provoca que el uso de las TIC en el entorno sanitario sea muy heterogéneo, tanto en lo referente al nivel de implantación como a las técnicas utilizadas. En el mejor de los casos, dentro de un dominio administrativo, es decir, en el ámbito en el que cierta administración ejerce sus competencias, se elabora una estrategia bien estudiada de desarrollo e implantación de sistemas de información sanitarios. Sin embargo, dado el modelo asistencial, también descentralizado, la atención al paciente suele recaer en manos de profesionales que pertenecen a distintos dominios administrativos, por lo que para tener una visión integral de la información que puede ser de interés en su atención sería necesaria la colaboración entre los sistemas implantados en distintos dominios.

Por tanto la integración de sistemas es una necesidad que surge, fundamentalmente, cuando se considera al paciente como el eje de la actividad sanitaria debido a que su atención es descentralizada. La dificultad que entraña la integración está causada por el modelo administrativo en este contexto, que provoca la heterogeneidad en soluciones estratégicas y tecnológicas. Haciendo esta consideración resulta especialmente interesante analizar, de forma breve, el papel que la información relativa a un paciente desempeña en la actividad asistencial y en el resto de actividades de la organización sanitaria.

1.1.2. La Historia Clínica

La historia clínica (HC) es el conjunto de documentos que surgen de la relación entre una persona, el paciente, y los agentes encargados de su atención sanitaria. Debe reflejar todo el conocimiento relativo al estado de salud del individuo y su situación como paciente, incluyendo juicios clínicos, proceso evolutivo, tratamiento, procedimientos realizados, recuperación, informes, consentimientos informados y cualquier otro documento que sea necesario o conveniente en la atención al sujeto.

Se confecciona a lo largo de la vida de una persona, actualizándose con los documentos que correspondan a cada proceso asistencial acontecido. Se origina con el primer episodio de enfermedad o control de salud del paciente y se termina con el último y más reciente. Si se considera que, por lo general, el primer contacto de un individuo con el entorno sanitario se produce en su nacimiento, la HC puede empezar a crearse desde ese mismo momento.

Los usos y beneficios de la HC son muy numerosos y entre ellos se puede destacar:

- La salud de una persona se ve afectada por numerosos parámetros y es difícil, al comenzar un proceso asistencial, recopilar todos en la primera entrevista (anamnesis) con el paciente. La historia clínica contribuye a mejorar la calidad asistencial ya que aporta una visión de conjunto del paciente antes incluso de entrevistarse con él. Esto permite al facultativo realizar un primer examen y evaluación mucho más eficientes que si tuviera que empezar de cero.
- Durante el resto del proceso asistencial la historia clínica recoge todos los informes generados y permite al facultativo tener, en cualquier instante, una visión clara de lo acontecido durante el mismo, lo que le facilitará tomar las decisiones pertinentes.

- Almacenar el registro completo de la atención prestada al paciente durante cualquier proceso asistencial posee una gran trascendencia como documento legal y puede ser importante cuando se planteen aspectos jurídico-legales relacionados con la atención de salud.
- Otra gran aportación es la contribución a la investigación, docencia y formación continuada, tanto en grado como postgrado, a través de sus revisiones o sesiones clínicas.
- Cada día se trabaja en mayor medida con medicina basada en la evidencia, donde la revisión de casos clínicos es la herramienta diaria del personal sanitario y esto sólo tiene utilidad si la HC ha sido bien ejecutada.
- La epidemiología estudia la salud en poblaciones humanas determinadas relacionándola con el medio que le rodea. Se considera la ciencia básica para la medicina preventiva y una fuente de información fundamental para la formulación de políticas de salud pública. Los estudios epidemiológicos se suelen apoyar en la revisión de una ingente cantidad de historias clínicas, procedimiento que puede resultar costoso, y sólo si las HCs se han realizado correctamente e incluyen la información necesaria, pueden utilizarse.
- La HC es también una fuente importante de conocimiento para la planificación, organización, gestión y evaluación sanitaria de la actividad asistencial de los diferentes servicios y unidades del hospital.

Así pues cabe destacar que las ventajas de la historia clínica no están solamente relacionadas con la asistencia al paciente al que se refiere, sino que la información recogida en ella puede ser utilizada en otras actividades dentro de la organización. Este es otro de los motivos que van a propiciar la necesidad de integración de sistemas que fueron creados para distintas aplicaciones pero que pueden beneficiarse unos de otros.

1.1.3. La Historia Clínica Electrónica

El uso de la HC se puede ver enriquecido si mejoran las técnicas de creación, mantenimiento y gestión de la historia clínica [2,3]. Las TIC suponen una revolución en este sentido [1]. Hasta hace poco, e incluso hoy en muchas ocasiones, la HC se materializaba en una colección de documentos en formato papel que se almacenaban en los centros de documentación de los hospitales, centros de salud o consultas.

Cada vez es más frecuente encontrar esta información en formato electrónico y que los profesionales sanitarios estén familiarizados con herramientas de gestión que facilitan todo el tratamiento de la misma. Como la información incluida en la HC puede ser de naturaleza muy diversa, no sería arriesgado aventurar que si se consiguiera almacenar toda la historia clínica en formato electrónico ésta se conformaría de numerosos registros multimedia, pudiendo incluir informes clínicos, análisis de laboratorio, imágenes, etc.

Esta irrupción de las nuevas tecnologías en el entorno sanitario, aunque va aumentando poco a poco, no es tan rápida como en otros escenarios a causa de una serie de barreras que se presentan en el contexto sanitario. Algunas de éstas son [5]:

- La escasez de equipos interdisciplinares. El coste de introducir expertos en tecnologías de la información en equipos médicos suele supeditarse a otros de índole asistencial más directa.
- La falta de formación específica en el dominio sanitario, por parte del personal técnico, y en el de manejo de nuevas tecnologías, por parte del personal sanitario. Esta "diferencia de lenguajes" dificulta el entendimiento entre los agentes necesariamente implicados en el desarrollo de una HC Electrónica (HCE).
- La propia naturaleza de la asistencia sanitaria, que es una actividad muy ligada a relaciones humanas y a procesos empíricos, hace que las prácticas y procedimientos sean difíciles de codificar en lenguaje "computacional". Los mecanismos de normalización y formalización de la información clínica son complejos.
- La variedad de agentes involucrados en actividades sanitarias aumenta la dificultad de normalizar una tecnología y de que ésta sea aceptada por todos los grupos implicados: facultativos y demás personal sanitario, técnicos, personal no sanitario, pacientes, administración, etc.

Revisar la situación actual de la HC requiere atención en tres puntos principales:

- ¿Cómo maneja el personal sanitario la HC?
- ¿Cuál es la penetración tecnológica actual?
- ¿Qué papel desempeña la HC en el conjunto de labores asistenciales?

Se analizan a continuación estas ideas y se presenta el escenario deseado para la integración de la HCE en la organización sanitaria y los problemas que la dificultan. Los esfuerzos realizados en esta Tesis Doctoral están encaminados a contribuir, en la medida de lo posible, a la consecución de este escenario final.

1.1.4. La gestión de la historia clínica

Debido a la creciente especialización de la Medicina y al modelo actual de sociedad, en la que cada vez es más acusada la movilidad de la población, los procesos asistenciales a un paciente suelen recaer en manos de distintos facultativos y equipos médicos, que podríamos denominar "proveedores de servicios sanitarios". Esto provoca que la historia clínica deba ser actualizada, mantenida y gestionada por numerosos profesionales que, en la mayoría de los casos, ni siquiera tienen contacto directo entre ellos e incluso pueden no pertenecer a la misma administración.

Esta situación da lugar a una fragmentación de la HC de una persona en HCs particulares por cada proveedor. Éstas se abren cuando el paciente realiza la primera visita, sin tener prácticamente ninguna información de HCs previas, salvo quizá por la emisión de informes cuando el paciente es remitido desde otro proveedor que le atendió previamente.

Recuperar todo el "conocimiento perdido" sobre el paciente supone un coste adicional en tiempo de entrevistas al paciente y/o familiares e incluso en la realización de procedimientos adicionales. Es más, parte del conocimiento almacenado en historias clínicas que se han elaborado de forma metódica y ejemplar por diversos profesionales será irrecuperable en consultas posteriores atendidas por otros proveedores.

Es evidente por tanto que utilizar mecanismos que permitan mantener una HC única para cada individuo y facilitar la consulta, modificación y gestión de la misma por los profesionales implicados en la atención del paciente, aportaría grandes ventajas asistenciales y supondría un considerable ahorro de recursos.

1.1.5. Tecnología en la historia clínica

Las TIC pueden aportar soluciones muy eficientes para la gestión de una HCE [6] única para cada paciente. Se han desarrollado numerosos sistemas de información clínica que permiten la gestión de la HCE en un determinado dominio y que pueden facilitar el intercambio o la gestión remota de esta información en formato electrónico.

Esta panacea se ve limitada por el modelo administrativo que sigue la atención sanitaria [7,8]. Las soluciones tecnológicas utilizadas para la gestión de una HCE son muy numerosas ya que se desarrollan sistemas adaptados a un dominio o administración concreta. En el mejor de los casos los sistemas están implantados por una administración de ámbito regional o nacional, lo que proporcionaría un escenario para conseguir la HCE única muy favorable. Sin embargo, la realidad es que los sistemas de tal cobertura se están desarrollando actualmente [9] y que los ya implantados y en funcionamiento cubren dominios de ámbito mucho más restringido, que pueden ir desde un hospital a un servicio concreto del mismo. De manera que estos sistemas suelen estar diseñados muy "a medida" y cubren las necesidades concretas de cada caso. Cada sistema utiliza sus propios modelos de información, semántica, lenguajes, plataformas, etc, dando lugar a un escenario de sistemas heterogéneos y en principio incompatibles.

Si toda la asistencia sanitaria que un determinado individuo necesita fuera cubierta dentro de uno de estos dominios todo estaría solucionado. Sin embargo cada vez es más frecuente que los equipos médicos que atienden a una persona a lo largo de su vida pertenezcan a distintas administraciones, lo que obliga a buscar soluciones a la integración de estos sistemas de información heterogéneos para conseguir gestionar una HCE única. Esta necesidad es patente incluso aunque se desarrollen sistemas de ámbito regional o nacional ya que, por un lado, es necesario respetar los sistemas implantados por administraciones de nivel inferior y por otro puede llegar el momento de la integración en un sistema de HCE única de nivel administrativo superior, por ejemplo dentro de la Unión Europea.

Es imprescindible mantener la autonomía de cada uno de los dominios implicados, que no pueden verse alterados por el hecho de pertenecer ahora a uno más amplio, y que deben seguir teniendo el control de su propia información. Este nuevo modelo se conoce como "Sistema Federado" [10] ya que se mantiene la autonomía de cada sistema de la federación pero se tiene un sistema global más completo y complejo.

1.1.6. La historia clínica en el conjunto de la actividad sanitaria

En una organización sanitaria que incluya múltiples centros, unidades e individuos, diferentes tipos de actores necesitarán compartir información y colaborar de acuerdo a determinados procesos y flujos de trabajo. Aunque los servicios finales ofrecidos por una organización sanitaria están principalmente centrados en la atención al paciente

dentro de la organización es necesario también realizar otras tareas, como investigación, formación, administración o gestión. De manera que en la labor asistencial se gestiona fundamentalmente información clínica pero hay otros muchos factores y tareas involucrados en la organización, manejándose también información financiera, de seguridad, de identificación y de naturalezas muy diversas.

La HC está centrada en la persona, ya que es el conjunto de documentos multiformato de carácter sanitario relativos a un individuo determinado, y es una pieza clave en la asistencia personal al paciente. En las actividades puramente asistenciales la consulta y modificación del historial clínico del paciente es crucial, aunque sólo representa una parte del conjunto de tareas que se llevan a cabo durante estas actividades y en las que no siempre se gestiona información estrictamente perteneciente a la historia clínica del paciente. El análisis de la información almacenada en la HCE y de las relaciones de ésta con otro tipo de información externa, por ejemplo genómica, proteómica o celular, es una fuente incalculable de conocimiento clínico, tanto relativo a un paciente concreto como referido a toda la población [11-14]. Este conocimiento puede ser explotado de muy diversos modos y puede suponer una mejora considerable de la calidad asistencial.

Por otro lado, la información almacenada en sistemas que se crearon con el propósito de gestionar la historia clínica puede ser de gran interés para el resto de actividades de la organización, no directamente asistenciales o centradas en un paciente, y puede ser la fuente de importantes parámetros de interés en la medida de la calidad de la organización.

Así pues es necesario tener en cuenta que la HCE es una pieza más dentro del conjunto de actividades sanitarias, ya sean asistenciales, científicas, administrativas o de cualquier otra naturaleza, y que pueden existir vínculos entre la información que se suele considerar parte de la HC de un sujeto y el resto de información manejada en el seno de la organización sanitaria. El análisis exhaustivo y la explotación eficiente de estas relaciones serán la base de la generación de conocimiento clínico, fundamental para el desarrollo de la Medicina del futuro.

Es de destacar que, dada la importancia de la gestión de la HC en la actividad asistencial, se han desarrollado gran variedad de sistemas independientes y heterogéneos. Esta situación ha provocado la necesidad de integración de dichos sistemas y, teniendo en cuenta las dificultades presentadas, gran parte de los esfuerzos de normalización en materia sanitaria se han centrado precisamente en la

estandarización de modelos para la HCE. Es un reto importante la integración de estos modelos con los que se desarrollen en el futuro para la materialización de una arquitectura de sistemas sanitarios. Al mismo tiempo los importantes avances realizados en el ámbito de sistemas de HCE pueden suponer una base importante en el desarrollo de estrategias de integración de sistemas en el seno de la organización sanitaria, facilitando la explotación eficiente del conocimiento almacenado en el conjunto de sistemas integrados en la arquitectura de la organización.

1.2. Motivación

1.2.1. Arquitectura objetivo

La HCE no debe entenderse como un fin en sí misma, sino que debe considerarse un pilar fundamental para facilitar la gestión eficaz del conocimiento sobre el estado de salud del paciente. De modo que si existe una estrategia de diseño e implantación de una arquitectura sanitaria dentro del dominio en el que se va a desarrollar un sistema gestión de HCE, un diseño eficiente del mismo debe considerar desde el primer momento que los componentes y aplicaciones para la gestión de la HCE se integrarán en esta arquitectura sanitaria compleja. Sobre ella se podrán desarrollar aplicaciones que cubran las necesidades, presentes y futuras, de los usuarios durante el desempeño de todas las actividades que tengan lugar dentro de la organización sanitaria [6,8]. Por supuesto será imprescindible que el modelo de arquitectura utilizado en la organización considere los componentes relacionados con esta importante actividad de gestión de HCE.

El diseño y desarrollo de nuevos sistemas de HCE y de métodos y técnicas para conseguir la HCE única no debe perder nunca de vista este hecho. Desde el principio la metodología de diseño del sistema de historia clínica debe considerar el paradigma de computación distribuida [15], es decir debe consistir en una serie de agentes con funcionalidades muy bien definidas que van a componerse para ofrecer prestaciones más complejas. Esto por un lado tiene la ventaja de poder aprovechar los componentes que cubran parte de las necesidades de este sistema de información y que ya existan en la arquitectura sanitaria y por otro facilita la incorporación del sistema de gestión de historia clínica dentro de la federación sanitaria.

Sin embargo no se puede obviar que actualmente existen numerosos sistemas de HCE en explotación que no han sido desarrollados conforme a este modelo. De forma que los

mecanismos de integración previstos en la organización deben facilitar la entrada de estos sistemas tan solo desplegando sencillas capas de intermediadores y sin obligarles a realizar cambios en el funcionamiento local. Así se permitirá que los sistemas de información que se integren en ella, que pueden pertenecer a distintas administraciones, sigan funcionando de forma autónoma pero que se incluyan en un entorno federado que proporcionará prestaciones más complejas.

La arquitectura facilitará la recuperación del conocimiento que un usuario autorizado concreto, en un contexto determinado y para la actividad que está desarrollando en ese momento, necesite. El caso de uso más evidente es la consulta, por parte de un facultativo, de la información clínica relativa a un individuo que le resulte de interés en un momento determinado, es decir la HCE única contextualizada, pero el sistema debe ser suficientemente modular, estructurado, flexible y escalable para facilitar el desarrollo de aplicaciones que cubran casos de uso más complejos.

De este modo el diseño de sistemas de gestión de la HCE debe enfrentarse como el diseño de elementos dentro de esta arquitectura de sistemas sanitarios. Este enfoque permite una alta reutilización de los componentes desarrollados en distintas aplicaciones relacionadas con las numerosas actividades de usuario y facilita la integración de sistemas siguiendo el paradigma de federación.

1.2.2. El problema de la federación de bases de conocimiento

En el campo de los sistemas de información aún queda mucho por hacer para conseguir que la integración de sistemas heterogéneos sea una tarea fácil. En esta sociedad inmersa en las TIC que nos ha tocado vivir se desarrollan numerosos sistemas de información con aplicaciones y tecnologías muy diversas. Sin embargo muchos de estos sistemas almacenan conocimiento relativo al mismo dominio y la integración de los mismos podría dar lugar a bases de conocimiento muy extensas y con infinidad de aplicaciones.

Los problemas de integración se plantean en tres niveles distintos [6,7,16-22]:

- **Nivel semántico:** Los modelos utilizados en cada base de conocimiento son distintos porque se han desarrollado de forma independiente. Sin embargo, si se refieren a un dominio común, habrá muchos conceptos que coincidirán en mayor o menor medida, aunque la representación sea distinta en cada uno. Será necesario encontrar un modelo semántico común que represente el conocimiento

almacenado en la federación de forma que todos lo entiendan de la misma forma: una ontología de integración. Esto evitará cualquier confusión semántica y facilitará la posterior automatización de tareas en este nivel. Este modelo debe ser suficientemente amplio para dar cabida a todos los conceptos del domino y fácil de ampliar y modificar en caso de ser necesario, hecho muy frecuente en dominios que cambian continuamente, como es el caso del sanitario. Sin embargo, es imprescindible mantener la autonomía de cada sistema individual, que debe seguir funcionando tal y como lo hacía antes de pertenecer a la federación. Por tanto no se debe modificar el esquema de datos interno a cada sistema sino encontrar las equivalencias del esquema de cada uno a la ontología común de la federación.

- **Nivel funcional:** Una vez puestos de acuerdo en la representación semántica del dominio será necesario saber cómo se maneja la información almacenada en cada sistema; es decir, habrá que definir un modelo funcional de las interfaces de cada sistema. En el nivel de aplicación el usuario necesita una vista homogénea del conocimiento almacenado en la federación. No se realizará una centralización de la información y se permitirá que cada sistema continúe trabajando como lo hacía antes, ocultando toda la complejidad de la integración al usuario. Así, las aplicaciones harán uso de las interfaces expuestas por cada sistema participante que permiten acceder a su información. Por supuesto la descripción del conocimiento gestionado debe estar basada en la ontología común. Si las interfaces de acceso a cada sistema se normalizan, es decir, si siguen modelos bien conocidos, el desarrollo de nuevos componentes que reutilicen estas interfaces es relativamente sencillo. La mayor complejidad es lograr el desarrollo de métodos que faciliten la integración de interfaces no normalizadas.

- **Nivel de instancia:** Para tener un vista global del conocimiento almacenado en todo el sistema sobre una entidad real determinada será necesario recuperarlo de los distintos sistemas de la federación que almacenen conocimiento relativo a esta entidad. Entonces dos nuevos problemas aparecen. En primer lugar es necesario identificar qué instancia dentro de cada sistema corresponde a la entidad real: éste es el problema de la identificación de identidad. En segundo lugar puede ser necesario resolver conflictos en los valores de los atributos de

las distintas instancias de la entidad, ya que el mismo atributo puede tener valores que no concuerden aunque realmente correspondan a la misma entidad. Como los sistemas van a seguir funcionando de forma autónoma, aunque se resuelvan los problemas en los niveles semántico y funcional los problemas a nivel de entidad pueden ser frecuentes.

La problemática expuesta se corresponde plenamente con la que se presenta al integrar distintos sistemas de HCE cuando el conocimiento sobre el estado de salud del paciente ha de ser gestionado de forma integrada. Y también en el caso más complejo de incorporación de dichos sistemas de HCE al conjunto de sistemas de información de una organización sanitaria.

1.3. Objetivos

El objetivo fundamental de esta Tesis Doctoral será proponer extensiones al modelo de arquitectura sanitaria que actualmente se está desarrollando en el contexto europeo [8] para facilitar la integración de sistemas siguiendo el modelo federado, mejorando la flexibilidad y escalabilidad de la misma. Estas extensiones deben considerar el problema de integración en su conjunto, proponiéndose métodos que ayuden a la resolución de los conflictos funcionales, semánticos y de instancia, especialmente para la incorporación de sistemas que han sido desarrollados siguiendo distintas normas o de forma completamente propietaria. Se analizarán especialmente los requisitos de esta arquitectura para cubrir las necesidades de gestión de la HCE facilitando la integración de estos sistemas con el resto de los de la organización.

Para ello se analizan los requisitos fundamentales de una arquitectura federada de sistemas sanitarios y los problemas de diseño que plantea. Se investiga, en particular, sobre el papel que los modelos de historia clínica ya desarrollados deben desempeñar dentro de esta arquitectura y se desarrollan modelos para llevar a cabo esta conciliación de forma eficiente. Se analiza además el contexto tecnológico actual y las facilidades y dificultades que plantea para las tareas de integración de sistemas siguiendo el paradigma de federación.

Como validación de los métodos propuestos y modelo de diseño, esta Tesis se centra en el servicio de identificación de personas o demográfico, comúnmente aceptado en trabajos de múltiples autores como fundamental en el dominio de la HCE y como el primer problema a resolver [23,24].

1.4. Estructura de la memoria

El presente documento se compone de las siguientes secciones y capítulos.

Sección 1. Introducción

Capítulo 1, Motivación y Objetivos: presenta la motivación de esta Tesis Doctoral y los objetivos planteados en la misma. Se describe brevemente la estructura de esta memoria.

Sección 2. Estado del Arte

Capítulo 2, La normalización en el contexto sanitario: en primer lugar se revisan los distintos esfuerzos de normalización en este dominio, desde estructuras de datos a servicios y arquitecturas. Se analiza qué organismos de normalización deben ser considerados para el desarrollo de sistemas de información sanitarios y cuál es el estado de la normalización, haciendo especial hincapié en los sistemas de historia clínica, los problemas de heterogeneidad de normas y la necesidad de integración.

Capítulo 3, Tecnologías de gestión semántica: se realiza una revisión de investigaciones y tecnologías con aplicación en el campo de la federación de sistemas. Se han revisado trabajos relativos a la reconciliación de entidades en una federación [17-20] y en lo que se refiere a las tecnologías que pueden resultar de interés en la resolución del problema de federación de sistemas se han estudiado algunas técnicas de integración como la arquitectura de agentes, de Servicios Web o la Grid. Se describe el concepto de ontología y de gestión semántica y cómo la combinación de ambas ideas está dando lugar al desarrollo de *middlewares*, o arquitecturas de integración, como la Web Semántica y la Grid Semántica. Finalmente se analizan los beneficios que estas técnicas pueden aportar al desarrollo de la arquitectura de sistemas sanitarios.

Capítulo 4, Métodos y modelos: en cuanto a los modelos en los que se puede basar la estrategia de implantación de una arquitectura sanitaria se analizan tres referentes clave: el modelo dual [25-28], la norma ODP [15] e HISA [8].

Capítulo 5, El servidor demográfico: por último, en lo que se refiere a las tareas de identificación de personas y gestión de sus datos personales, se estudiada en detalle el modelo de servicio de identificación de personas [29] propuesto por el grupo CORBAmed del OMG, analizando puntos fuertes y débiles, y se revisan trabajos previos en la identificación de personas en entornos federados [30-32], en especial los

aplicados al contexto sanitario. También se presentan en este capítulo algunos de los modelos de datos demográficos normalizados más importantes [33,34].

Sección 3. Resultados y Validación

Capítulo 6, Resultados. Extensiones a HISA: se plantean extensiones a HISA para conseguir una arquitectura más flexible y escalable dando como resultado un método de diseño y desarrollo de sistemas que facilita la integración de sistemas heterogéneos siguiendo el paradigma de federación.

Capítulo 7, Validación. Gestión demográfica: utilizando la metodología desarrollada se presenta una arquitectura para la identificación de personas y la gestión de datos personales en un entorno federado.

Sección 4. Conclusiones

Capítulo 8, Contribuciones originales: se presentan las contribuciones principales de esta Tesis Doctoral

Capítulo 9, Publicaciones: en este capítulo se presentan, brevemente, las publicaciones a las que ha dado lugar el desarrollo de esta Tesis.

Capítulo 10, Línea de investigación: se presenta la línea de investigación que la autora de esta Tesis seguirá a partir del momento de la obtención del título de doctor.

Capítulo 11, Bibliografía: donde se reflejan las principales referencias de esta Tesis Doctoral.

Capítulo 2. Normalización

2.1. Introducción

El objetivo principal de la normalización es facilitar la producción, manejo y uso de productos o servicios en el marco del libre mercado para satisfacer a clientes y suministradores.

La utilización de normas en el sector de las TIC sanitarias puede traer grandes beneficios entre los que son destacables [7,35]:

- Facilitar el interfuncionamiento entre sistemas y el intercambio de información de pacientes entre organizaciones sanitarias.
- Aumentar la eficiencia del mercado mejorando el entendimiento entre todas las partes a través de un marco técnico y de términos para el desarrollo e implementación de aplicaciones sanitarias.
- Proveer mecanismos para la seguridad, confidencialidad y cumplimiento de los requisitos legales.
- Establecer un conjunto de servicios interoperables, coordinados y abiertos basados en un negocio y una arquitectura comunes a varios dominios.

La adopción de estándares, aunque necesaria, puede implicar un coste significativo en cuanto a tiempo y dinero, que suele representar una barrera para algunas organizaciones [7], en especial las pequeñas y medianas empresas. Sin embargo, no hay que olvidar que los miembros de la Unión Europea tienen la obligación, según las directivas de la misma, de utilizar estándares europeos o globales en aquellas materias donde existan. Pero, hasta la fecha, hay pocos fabricantes que basen sus productos en estándares europeos: para vender dentro del contexto pan-Europeo la mayoría basan sus productos en los estándares americanos, ya que EEUU es la sede de la mayor parte de empresas que desarrollan productos en este ámbito.

La falta de implementaciones de la normativa europea está causada por varios factores entre los que se pueden destacar:

- Los mecanismos de desarrollo del Comité Europeo de Normalización (CEN) son demasiado lentos y burocráticos. Se carece de políticas de marketing y apoyo al desarrollo y en la mayoría de las ocasiones las normas no son de acceso gratuito.
- La no existencia de un perfil completo de estándares que permitan su aplicación en todo el contexto sanitario.
- La falta de productos ya desarrollados que sirvan de modelo.
- Los conflictos que a veces existen de las normas del CEN con las de otros organismos como DICOM, HL7 o IEEE.
- La escasez de mecanismos de validación de la interoperatividad cuando los estándares sí existen.

Estudiar a fondo el marco normativo de las TIC en sanidad es un ejercicio imprescindible a la hora de realizar cualquier investigación, desarrollo o elaborar un plan estratégico de implantación de estas tecnologías. El grupo denominado CEN/ISSS e-Health Focus Group se formó precisamente para preparar un informe que incluyera el estado del arte sobre los procesos de normalización actual y futura en el dominio sanitario, así como un análisis de la estrategia a seguir para la aplicación de las TIC en el ámbito sanitario, dentro siempre del contexto europeo. En el portal del CEN, para su comentario público, se publicó en marzo de 2005 este informe bajo la denominación **"Current and future standardization issues in the e-Health domain: Achieving interoperability"** [7]. Éste constituye un excelente documento de referencia a la hora de analizar el estado actual de normalización en tecnología sanitaria y las pautas a seguir en el futuro más inmediato en cuanto a normalización e implantación.

La necesidad de la elaboración de este informe y la creación del grupo adecuado para llevarlo a cabo surge a raíz de la declaración ministerial europea del 22 de Mayo de 2003. Esta declaración fue el fruto de la conferencia de 2003 sobre materia de e-Health organizada por la Comisión Europea y la presidencia griega del Consejo. Se entendía por e-Health el uso de tecnologías de información y comunicación actuales que ayudaran a cubrir las necesidades de ciudadanos, pacientes, profesionales y proveedores sanitarios y administraciones. Uno de los mensajes de esta declaración fue que los ministros apoyarían acciones para el desarrollo de estándares que permitieran la interoperatividad de sistemas y servicios, especialmente aquellas que exploraran las posibilidades del uso de aplicaciones de código abierto para lograr este objetivo.

La estrategia seguida para la realización de este informe fue dividir las tareas en 5 grupos principales (Work Packages):

- WP1: Identificación de prioridades de aplicación de las TIC en el entorno sanitario desde el punto de vista de estrategias, políticas o planes de administraciones nacionales.
- WP2: Revisión de políticas de normalización nacionales y europeas.
- WP3: Definición de prioridades de actuación desde el punto de vista de todas las partes implicadas.
- WP4: Revisión y clasificación de estándares y programas de trabajo existentes.
- WP5: Análisis del estado del arte y elaboración de recomendaciones.

2.2. Estrategias de normalización para el interfuncionamiento

Las conclusiones del informe mencionado [36] son tremendamente interesantes. Aunque no es posible hacer aquí un análisis profundo de este informe se destaca la discusión de aspectos relacionados con la interoperatividad, debido a la afinidad de dicho capítulo con la temática principal de este trabajo de tesis.

Se diferencian tres formas en las que los sistemas pueden cooperar. La más básica es el desarrollo de interfaces específicas entre cada sistema implicado. Cuando el número de sistemas crece esta tarea resulta prácticamente imposible. El método más complejo de cooperación es la integración, que consiste en conseguir que varios componentes trabajen juntos como uno solo. Generalmente un sistema de salud no trabajará siempre como una organización única y completamente integrada y la mayoría de sistemas implicados cooperarán sólo de forma temporal según el proceso acometido en cada momento. Aunque existirán acuerdos permanentes entre las partes implicadas, cada uno seguirá teniendo sus restricciones locales que prevalecerán sobre los del sistema integrado, esto es lo que podemos denominar una federación. El interfuncionamiento consiste en definir mecanismos para que socios heterogéneos colaboren para proporcionar determinado servicio.

2.3. Organismos oficiales de normalización

A continuación se presenta una breve introducción a los principales organismos de normalización que, o bien se dedican al ámbito sanitario, o bien tienen líneas de actuación centradas en este tema.

2.3.1. CEN/TC251 "Health Informatics"

El Comité Europeo de Normalización (CEN) [37], desarrolla trabajos de normalización que cubren todos los sectores técnicos a excepción del campo electrotécnico, que es competencia del Comité Europeo de Normalización Electrotécnica (CENELEC) [38]. Su papel es crear normas europeas que fomenten la competitividad de la industria europea a nivel mundial y ayuden a mejorar el mercado interior europeo. Sus principales objetivos son preparar nuevas normas europeas o documentos de armonización sobre aquellos temas en los que no existen normas internacionales o nacionales y promover la implantación en Europa de las normas desarrolladas por ISO. Genera varios documentos normativos, entre ellos normas europeas (EN), que son de obligado cumplimiento por los miembros adoptándose como normas nacionales, y normas experimentales europeas (ENV), que son documentos elaborados por los miembros para su aprobación provisional en aquellos campos donde exista un alto grado de innovación tecnológica o una urgente necesidad de orientación.

Dentro del CEN, el comité técnico 251 (Health Informatics) [36] es el organismo encargado del desarrollo de estándares para tecnologías de la información relacionadas con Medicina. Se creó en 1990 y desde entonces ha publicado alrededor de 50 documentos técnicos entre los que se encuentran prenormas, normas (que resultan de la revisión de los anteriores) e informes. Las tareas se reparten entre 4 grupos de trabajo (Works groups, WG) que se centran en:

- **WG-I, Modelos de información:** El objetivo es facilitar la comunicación entre sistemas de información independientes. Entre otras tareas se definen normas para representar y almacenar la información y mecanismos y reglas para compartirla e intercambiarla. Cualquier estándar desarrollado por este grupo debe asegurar que la información intercambiada no presente ninguna ambigüedad, para asegurar un alto nivel de calidad en los cuidados del paciente. Este grupo de trabajo colabora estrechamente con el resto para asegurar que se cumplen todos los requisitos establecidos por los mismos.

- **WG-II, Terminologías y representación del conocimiento:** Sus objetivos principales son la organización semántica de la información sanitaria para que pueda ser utilizada en distintos dominios, así como ofrecer mecanismos y criterios para conseguir una consistencia en la terminología utilizada dentro del TC251.
- **WG-III, Seguridad, confidencialidad y calidad:** Las actuales legislaciones, tanto a nivel europeo como nacional, suelen enfatizar la importancia de utilizar mecanismos que aseguren la calidad, confidencialidad y seguridad, especialmente cuando se trata con información tan delicada como la relativa a la salud. Se suele definir la seguridad de sistemas de información como la prevención de problemas de confidencialidad, integridad y disponibilidad de dichos sistemas, y especialmente cuando la información es relativa a la salud de las personas es imprescindible asegurar la privacidad de cada individuo. Este grupo de trabajo se encarga de asegurar que existen estos mecanismos. Casi todos los documentos generados se apoyan en las recomendaciones del Consejo Europeo y en las directivas de protección de datos de la Unión Europea. Los objetivos principales son desarrollar perfiles de protección en distintos sectores o ámbitos de aplicación y protocolos para ofrecer servicios de seguridad, así como evaluar, certificar y acreditar los productos y sistemas que se desarrollen en el mercado.
- **WG-IV, Tecnologías para la interoperatividad:** El objetivo de este grupo de trabajo es desarrollar estándares que faciliten el interfuncionamiento de dispositivos y sistemas de información en entornos médicos. Cubre tres áreas principales: intercomunicación de datos entre dispositivos y sistemas de información, integración de datos para la presentación multimedia y comunicación de datos entre departamentos y otros usuarios del sector.

Se destacan a continuación, dada la gran influencia de las mismas en esta Tesis Doctoral, varias prenormas, es decir que todavía están en proceso de elaboración y aceptación, que fueron creadas con el propósito de lograr interoperatividad e integración de sistemas de información clínicos:

La prenorma prEN 13606: Comunicación de historia clínica electrónica [27,28,39-41], elaborada a partir de la norma experimental ENV13606 de 2000.

- Parte 1: Modelo de referencia
- Parte 2: Modelo de arquetipo
- Parte 3: Metodología para el modelado de dominios clínicos
- Parte 4: Seguridad
- Parte 5: Modelos de intercambio

La prenorma prEN 14822: Componentes de información de propósito general (GPICS) [34,42,43]

- Parte 1: Visión general
- Parte 2: No clínicos
- Parte 3: Clínicos
- Parte 4: Cabeceras de mensajes

La prenorma prEN 12967: Arquitectura de servicios (HISA) [8,44,45]. Elaborado a partir del estándar experimental ENV 12967 de 1998.

- Parte 1: Vista de empresa
- Parte 2: Vista de información
- Parte 3: Vista de computación

Un proveedor que quiera desarrollar cualquier sistema para incorporarlo a la arquitectura de una organización sanitaria y busque el interfuncionamiento dentro del marco de trabajo europeo debe considerar estos estándares, pero por supuesto también otros que sean específicos del tipo de información que el sistema en concreto pretenda gestionar.

2.3.2. ISO/TC215 "Health Informatics"

El comité técnico 215 de informática sanitaria de la ISO [46] fue creado en 1998. Su objetivo es la definición de estándares relativos a las TIC en el entorno sanitario para conseguir la compatibilidad e interoperatividad entre sistemas independientes. Además quiere asegurar la compatibilidad de los datos para que puedan ser utilizados con propósitos estadísticos y reducir los esfuerzos duplicados y redundancias.

ISO es un organismo de carácter internacional, por lo que desde el primer momento se ha luchado por una colaboración estrecha entre organismos, intentando consensuar normas americanas y europeas, lo que en muchas ocasiones resulta complicado. Participan 25 países más 14 que sólo son observadores. Tiene alianzas con varias organizaciones como el CEN, HL7 o el W3C, y de hecho algunos de los grupos de trabajo más importantes de la ISO están dirigidos por miembros del CEN o del HL7. Los grupos de trabajo del comité técnico 215 son:

- WG1: Coordinación para el modelado y almacenamiento de datos sanitarios.
- WG2: Mensajes y comunicación.
- WG3: Representación de conceptos de salud.
- WG4: Seguridad.
- WG5: Tarjetas sanitarias.

2.3.3. Health Level Seven: HL7

El HL7 [47] es un organismo norteamericano que fue fundado en 1987 por varios proveedores de software para la industria sanitaria. Su finalidad es crear estándares para el intercambio, gestión e integración de datos que sirvan de apoyo al cuidado clínico del paciente y a la gestión, prestación y evaluación de servicios sanitarios. Su estrategia se ha centrado en la especificación de mensajes con formato consensuado que facilitaran la interoperatividad de los sistemas de información hospitalarios (Hospital Information System, HIS). En el año 1994 el HL7 fue acreditado por el ANSI (American National Standards Institute) como un organismo desarrollador de estándares (Standards Development Organization, SDO), lo que significa que las especificaciones aprobadas en el HL7 pueden pasar al proceso oficial de normalización como un estándar nacional americano.

La versión 1.0 de la norma, que especifica el formato de los mensajes a intercambiar entre HIS, fue aprobada en 1987 y seguida de la versión 2.0 en 1998. Actualmente esta versión es la utilizada en muchos HIS americanos y en algunos países europeos. En el año 2000 se publica un conjunto de especificaciones que provee una serie de estructuras de información, basadas en XML, denominado Arquitectura de documentos clínicos (CDA, Clinical Document Architecture). En futuras revisiones el número de estructuras de información irá ampliándose. Estas estructuras están pensadas para ser

intercambiadas utilizando los mensajes definidos en HL7 y están incluidas en el modelo de información de HL7 (RIM, Referente Information Model). En la actualidad se está ultimando el desarrollo de la versión 3 de mensajes para la interoperatividad, donde se especifica una metodología para el desarrollo de mensajes basados en el modelo de referencia con la intención de facilitar una implementación más consistente que en la versión 2.

A la hora de elaborar cualquier estrategia de integración o implantación de las TIC en el entorno sanitario será imprescindible considerar en algún momento los estándares desarrollados por HL7, ya que los proveedores de sistemas más importantes (Siemens, GE Medical Systems, HBOC-McKesson, IBM, Oracle, Microsoft, Philips…), desarrollan sus productos conforme a estos estándares y por tanto son utilizados en numerosos sistemas ya desplegados. Sin embargo, los mecanismos de HL7 para la interoperatividad están únicamente basados en la definición de mensajes y estructuras de información y esta aproximación al problema hace que no sea sencillo integrar sistemas no desarrollados conforme a esta norma, suponiendo un coste excesivo y en ocasiones imposible de asumir.

2.4. Otros organismos de normalización

Los organismos mencionados a continuación desarrollan estándares que vienen siendo utilizados ampliamente por diversos fabricantes y que además influyen en la elaboración de normas por parte de los organismos oficiales.

2.4.1. ACR-NEMA (DICOM)

En 1983 se crea una asociación entre dos importantes organismos americanos, el American Collage of Radiology (ACR) y la National Electrical Manufacturers Association (NEMA) [48], cuyo objetivo primordial es desarrollar normas para el tratamiento y transmisión de imágenes médicas y su información asociada. Los primeros resultados se publicaron en 1985 y hubo una versión posterior en 1988. Sin embargo, el rápido avance de las TIC provoca que pronto queden obsoletos y se decide rediseñar el estándar para integrarlo con los nuevos estándares en redes y comunicaciones, como se muestra en la figura 1, dando como resultado DICOM (Digital Imaging and COmmunications in Medicine) [48-50] que se publica en 1993.

Aunque su origen es americano actualmente se utiliza, de forma total o parcial, en casi todo el mundo.

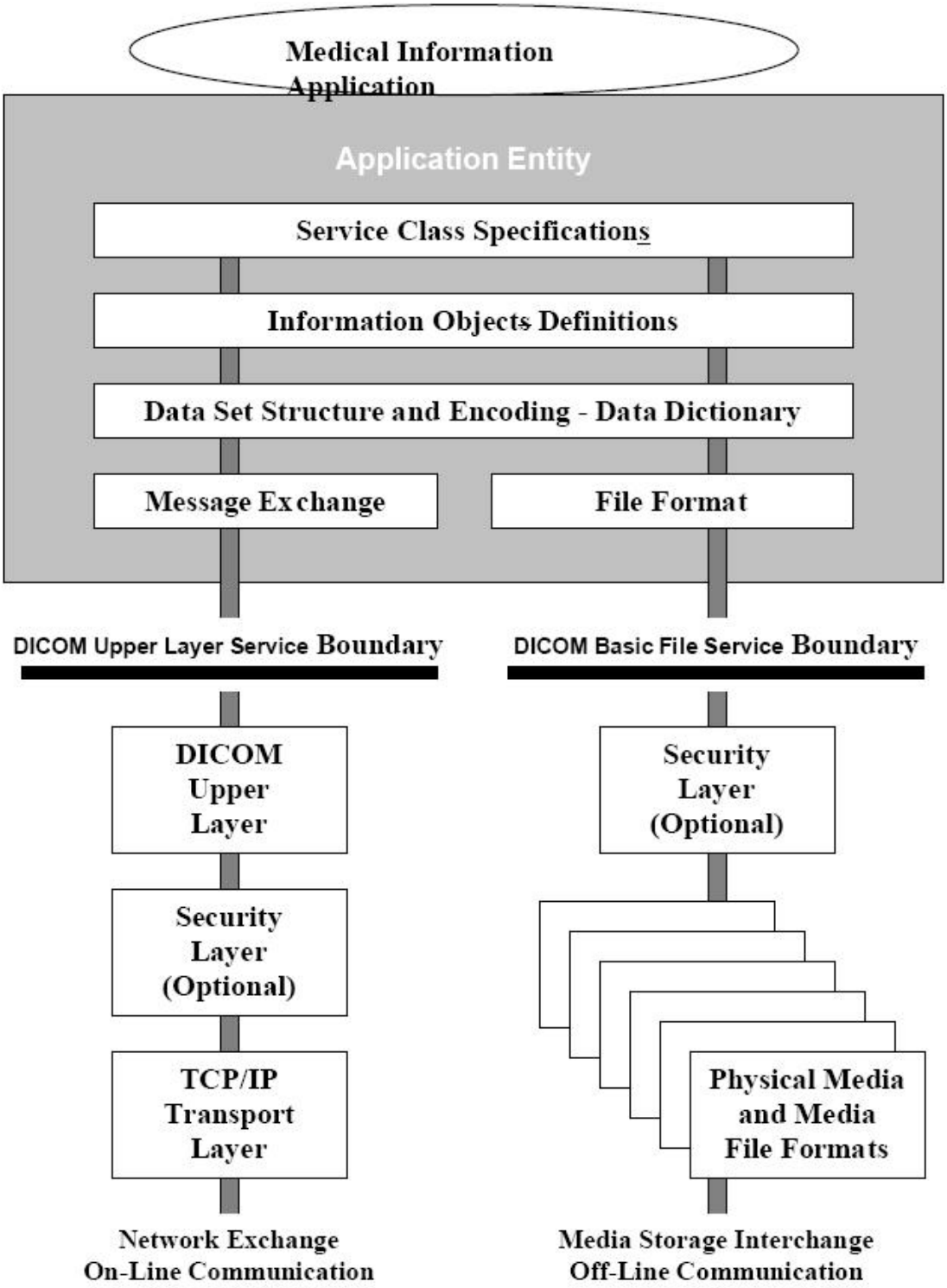


Figura 1. Modelo de comunicaciones en DICOM (extraído de [51])

Este estándar es mantenido y ampliado por un comité multidisciplinar internacional. Inicialmente estaba muy orientado a la radiología pero actualmente incluye otros

muchos tipos de imágenes y pruebas, como por ejemplo electrocardiogramas. DICOM está dividido en varias partes que se nombran como "PS 3.N-YYYY", donde "N" identifica la parte y "YYYY" el año de publicación. Cada parte es independiente de las demás y su estado de desarrollo puede ser distinto. En la última revisión realizada estas partes son:

- PS 3.1-2004: Introducción y generalidades.
- PS 3.2-2004: Conformidad.
- PS 3.3-2004: Definiciones de los objetos de información.
- PS 3.4-2004: Especificaciones de las clases de servicio.
- PS 3.5-2004: Estructura de los datos y codificación.
- PS 3.6-2004: Glosario de términos.
- PS 3.7-2004: Intercambio de mensajes.
- PS 3.8-2004: Soporte de red para el intercambio de mensajes.
- PS 3.10-2004: Medios de almacenamiento y formato de ficheros para el intercambio de datos.
- PS 3.11-2004: Perfiles de aplicaciones para medios de almacenamiento.
- PS 3.12-2004: Formatos y medios físicos para el intercambio de información.
- PS 3.14-2004: Funciones estándar para presentación en escala de grises.
- PS 3.15-2004: Perfiles de gestión del sistema y seguridad.
- PS 3.16-2004: Recursos para establecer equivalencias de contenido.
- PS 3.17-2004: Información explicativa.
- PS 3.18-2004: Acceso Web a objetos persistentes DICOM (WADO)

En cuanto a la comunicación entre sistemas, aunque DICOM está basado en la torre OSI se permiten implementaciones siguiendo otros modelos. De hecho, el uso de TCP/IP junto con las capas superiores de OSI está muy extendido en el desarrollo de aplicaciones que siguen el estándar DICOM.

En lo referente al almacenamiento de información, se definen distintas clases de servicio de almacenamiento (Creador, Lector y Actualizador de ficheros). Se especifica

cómo se debe almacenar la información en un soporte físico. Se utiliza una cabecera que contiene identificadores del tipo de información y de la sintaxis de transferencia que se debe utilizar.

Para terminar con este breve recorrido por el estándar DICOM sólo destacar que se define un modelo de información que representa la forma en que se estructura la información. Se establecen cuatro niveles en este modelo:

- **Nivel de Paciente:** Identificación e información sobre el paciente al que pertenece el estudio. Si consideramos toda la información relativa a un mismo paciente este nivel sería el más alto, ya que un mismo paciente podrá tener varios estudios.
- **Nivel de Estudio:** Un estudio será el resultado de una petición para un determinado tipo de examen sobre un paciente. Deberá mantener información de identificación del paciente así como otras referencias procedentes de otras fuentes, como podría ser el HIS.
- **Nivel de Serie:** Como resultado de un estudio se obtendría un conjunto de imágenes que se pueden organizar en series. En este nivel se recogerá información sobre el tipo de equipo que generó los resultados, la fecha y hora y otros detalles acerca del examen.
- **Nivel de Imagen:** Será el nivel inferior, cada imagen tiene información sobre su adquisición y sobre la posición del paciente, además de la imagen propiamente dicha. Según el equipo la imagen podría ser simple o una serie. El uso de estas últimas sólo es posible cuando la relación entre ellas puede ser descrita de forma simple.

2.4.2. The Institute of Electrical and Electronics Engineers (IEEE)

El IEEE lleva a cabo actividades de normalización en los EEUU a través de la organización subsidiaria Institute of Electrical and Engineers Standards Association (IEEE-SA), que desarrolla estándares para un amplio rango de industrias, incluida la biomedicina y atención sanitaria. Existen colaboraciones entre IEEE, CEN/TC251 e ISO/TC215. Los esfuerzos de normalización en el IEEE dentro del dominio sanitario son principalmente en el ámbito de la comunicación entre dispositivos médicos.

2.4.3. CORBAmed del OMG

El grupo de trabajo CORBAmed [52], perteneciente al Object Management Group (OMG), desarrolla una serie de normas centradas en la definición de servicios con propósitos específicos dentro de una arquitectura de sistemas sanitarios. Algunos de los servicios más importantes son:

- **De identificación de personas (Person Identification Service, PIDS)** [29]: da una serie de funcionalidades para gestionar de forma apropiada los identificadores de personas. Está pensado para facilitar la gestión de identificadores, y por tanto la localización de personas, no solamente dentro de un sistema particular sino también a través de distintos dominios.
- **De acceso a términos (Lexicon Query Service, LQS)** [53]: el objetivo de este servicio es especificar una serie de funciones de lectura para acceder al contenido de sistemas con terminología médica, que podría ir desde una simple lista con códigos hasta una base de datos dinámica con estructura mucho más compleja. La idea es dar una especificación que pueda ser utilizada para realizar interfaces con la mayoría de los sistemas de codificación médicos.
- **De acceso a observaciones clínicas (Clinical Observation Access Service, COAS) [54]**: se definen interfaces y estructuras de datos que facilitan el acceso a observaciones clínicas de cualquier tipo.
- **De acceso a imágenes clínicas (Clinical Image Access Service, CIAS)** [55]: se definen interfaces y estructuras de datos que facilitan el acceso a imágenes médicas y los metadatos relacionados con éstas. Se proporcionan mecanismos para transformar imágenes con distintos formatos, entre ellos el formato DICOM. Sin embargo, curiosamente, sólo está pensado para aplicaciones que no pretendan realizar una diagnosis a partir de estas imágenes.
- **De decisiones para el acceso a los recursos (Resource Access Decision, RAD)** [56]: las facilidades proporcionadas serán una serie de mecanismos para obtener decisiones de autorización y para administrar las políticas por las que se rigen estas decisiones. Más concretamente se proporcionan mecanismos que no se facilitaban con el servicio de seguridad de CORBA ya existente y que son imprescindibles en los entornos sanitarios.

Esta arquitectura de servicios ha servido de referente a diversos trabajos [57-60] y estándares, como el ya mencionado prEN 12967 (o HISA), que buscan solucionar el problema de la interoperatividad funcional. Como sucede en todos los grupos dentro del OMG, la normalización se basa en especificar los tipos de datos útiles en el servicio de interés y las interfaces de acceso a los mismos. Algunos de los servicios normalizados o en proceso de normalización se presentan en la figura 2.

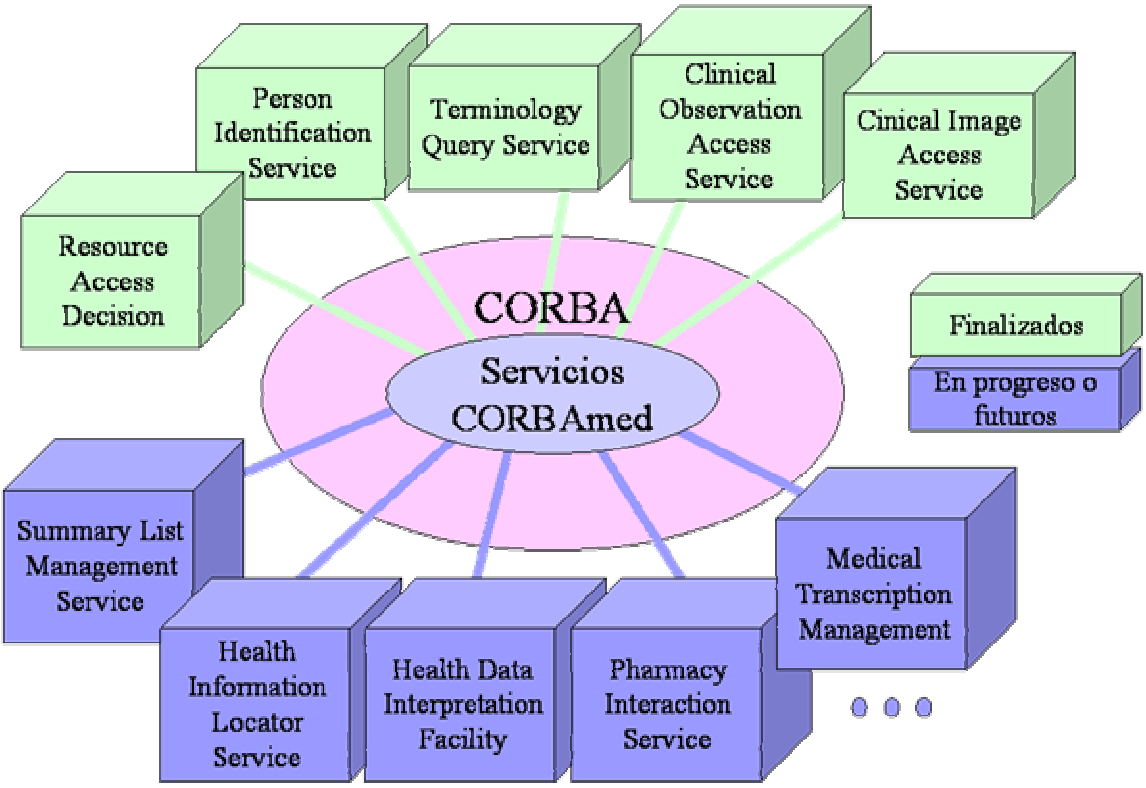


Figura 2. Algunos servicios definidos por CORBAmed

2.4.4. Estándares de terminología

Cabe destacar también los esfuerzos de varias organizaciones para desarrollar sistemas de clasificación y codificación que faciliten la resolución de conflictos semánticos, tan críticos en el intercambio de información sanitaria. Se enumeran a continuación algunos de estos sistemas de términos:

- Unified Medical Language System (UMLS) [61].
- International Classification of Diseases, 10th Edition (ICD-10) [62].
- The Systematized Nomenclature of Medicine (SNOMED) [63].
- Laboratory Observation Identifier, Names and Codes (LOINC) [63,64].

- International Union of Pure Applied Chemistry (IUPAC) Recommendations
- Multilingual glossary of technical and popular medical terms in nine European languages (EUGLOSS) [65].
- ISO 5218:1977 Information interchange- Representation of human sexes

2.5. Iniciativas de apoyo a estándares

El trabajo realizado por los grupos anteriores se ve notablemente influenciado por proyectos de investigación y por los esfuerzos de otros grupos de trabajo que, sin ser organismos de normalización, influyen notablemente en estos. Es necesario tenerlos en cuenta al hacer una revisión de las normas adoptadas en entornos sanitarios.

2.5.1. OpenEHR

Fundación que persigue el desarrollo de normas abiertas (Open) para la gestión de registros sanitarios electrónicos (Electronic Healthcare Record, EHR), es decir normas para el desarrollo de una HCE que facilite la interoperatividad. Su trabajo se centra en la especificación de un modelo de referencia [33,66-70], bastante similar al descrito en el prEN 13606, y un modelo de arquetipos [26,71] que añade restricciones sobre el anterior para conseguir conceptos más próximos al dominio clínico. Esta idea de arquetipos ha sido adoptada por el CEN en la última revisión del prEN 13606, lo que demuestra la influencia de estos grupos independientes en el desarrollo de estándares. Aunque la filosofía del OpenEHR también es facilitar la integración de sistemas de información clínicos, el nivel de desarrollo actual más bien permite un desarrollo local muy bien planificado, fácil de gestionar y extender, y da un marco para la definición de semántica clínica que facilite la interoperatividad. Sin embargo aún no están definidos los servicios implicados ni los flujos de trabajo de las actividades sanitarias.

2.5.2. Integrating the Healthcare Enterprise: IHE

El IHE es una iniciativa privada e independiente que resulta de la asociación entre fabricantes y sociedades profesionales [72]. El objetivo de este organismo es la especificación de perfiles de integración especializados basados en estándares ya existentes y en la definición de procedimientos para validar la interoperatividad. Los perfiles de integración se definen como la agrupación de actores, transacciones y vocabulario común para realizar una tarea típica de flujo de trabajo en una organización

sanitaria. Su trabajo parte de estándares estables para construir un conjunto completo de normas que cubran todo el espectro sanitario, pero a su vez realimentan con sus esfuerzos a los desarrolladores de estos estándares originales. Actualmente su objetivo está centrado en los flujos de trabajo dentro de un hospital, aunque se están desarrollando ya perfiles que incluyen también el intercambio de documentos entre distintos proveedores de servicios.

2.6. Conclusión

Este escenario tan heterogéneo de organismos de normalización, sectores implicados e intereses, constituye un marco en el que la adopción de un estándar dentro de una comunidad determinada es todavía complicada. Por otro lado los sistemas que ya se han desplegado o van a desplegarse próximamente no suelen cumplir normas, o si lo hacen pueden haber seleccionado aquellas que al final no serán adoptadas por la comunidad a la que pertenecen. Las normas de integración e interoperatividad que se desarrollen no pueden hacerlo a espaldas de esta situación ya establecida, es decir, que deben facilitar la integración de sistemas ya desarrollados con el menor coste posible.

Aunque en determinados momentos de esta Tesis Doctoral se hace referencia a normas, principalmente europeas, la metodología desarrollada no obliga a los sistemas a integrar a que las adopten en su estructura interna y tampoco está condicionada al uso de la normativa europea. Por otro lado la metodología empleada es extensible a la integración de la historia clínica con sistemas de información no clínicos, es decir, que se buscan mecanismos de integración total que son difíciles de conseguir con estándares completamente centrados en el terreno sanitario.

Capítulo 3. Técnicas de semántica

3.1. Introducción

Uno de los problemas que surgen en la integración de sistemas es el conflicto funcional, es decir, cómo colaboran los sistemas que se incluirán dentro de una federación. En este sentido, desde el punto de vista tecnológico, será necesario revisar las plataformas de integración existentes (*middleware*) y analizar las más emergentes, como los servicios Web o la Grid.

Otro de los requisitos en la federación de sistemas es resolver los conflictos semánticos entre distintos sistemas; es decir, llegar a un entendimiento común de la información compartida por los dominios que se pretenden integrar. La solución pasa por la creación de sistemas de términos y modelos de datos compartidos que dan lugar a lo que podría denominarse un esquema federado. Las cuestiones son si el esquema diseñado cubre los conceptos necesarios para las interacciones entre cualquier sistema que pretenda incorporarse a la federación, cuánto esfuerzo se requiere para traducir los conceptos locales en cada dominio a los federados, y si realmente el esquema diseñado servirá a largo plazo. Entender este esquema federado como un idioma dinámico, en el sentido de que puedan añadirse conceptos, sinónimos, etc, y donde se puedan manejar y extender eficazmente las relaciones entre conceptos, permitirá que las arquitecturas desarrolladas con esta filosofía no queden obsoletas cuando evolucione el dominio de interés y facilitará la incorporación de nuevos dominios. Así pues, al investigar sobre federación de sistemas resulta imprescindible analizar el estado del arte en técnicas de gestión semántica.

La necesidad de resolver ambos conflictos en cualquier proceso de federación y el reconocimiento de este hecho por la comunidad científica ha dado lugar a la aparición de técnicas que combinan estos dos campos para desarrollar plataformas de integración basadas en la gestión semántica como la Web Semántica o la GRID Semántica.

3.2. Plataformas de integración

3.2.1. Objetos distribuidos

Las plataformas de objetos distribuidos tienen su base conceptual en un paradigma bien asentado en los últimos años: la orientación a objetos. Los objetos distribuidos se pueden describir como "cápsulas" de información que residen en cualquier lugar de la red [73,74]. Los clientes pueden acceder a ellos a través de la invocación de métodos, siendo para este acceso transparentes el sistema operativo, el compilador y la plataforma hardware utilizada. Algunas de las ventajas que se pueden destacar al usar esta tecnología son:

- Facilita la creación de sistemas cliente/servidor flexibles.
- Los objetos son entidades autogestionadas.
- Facilita la interoperatividad a través de red.
- Los objetos pueden funcionar en distintas plataformas.

El OMG (Object Management Group) [75] fue fundado en el año 1989 por once compañías. Su intención era desarrollar especificaciones abiertas para el desarrollo de software orientado a objetos. La especificación fundamental del OMG es una plataforma de procesamiento distribuido orientada a objetos denominada CORBA (Common Object Request Broker Architecture). Ésta proporciona un marco de trabajo para la descripción y estandarización de sistemas abiertos de procesamiento distribuido. Las aplicaciones y servicios que se desarrollen siguiendo este paradigma permitirán poner en ejecución los distintos objetos en distintas máquinas comunicadas a través de una red, aunque pertenezcan a distintos fabricantes y tipos de sistemas.

El lenguaje de especificación de interfaces normalizado por el OMG es IDL (Interface Definition Language), que permite definir los objetos CORBA y sus relaciones. Así una definición en IDL declara un juego de funciones a las que puede acceder un cliente CORBA, los tipos de datos que se intercambian y las excepciones que pueden surgir.

CORBA ofrece grandes facilidades para conseguir la interoperatividad entre sistemas heterogéneos distribuidos ocultando la complejidad de las comunicaciones en red y ha sido ampliamente utilizado en entornos médicos [58,76,77] de modo que incluso dentro

del OMG se ha creado un grupo de trabajo especializado en servicios orientados a objetos para aplicaciones médicas (CORBAmed) [52].

3.2.2. Agentes Software

El paradigma multiagente propone una nueva concepción de la programación distribuida, en la que el comportamiento del sistema resulta de la compleja interacción entre entidades software denominadas agentes. El concepto de Agente o Agente Inteligente surge en la comunidad de Inteligencia Artificial Distribuida (Distributed Artificial Intelligence, DAI). La definición de agente no resulta fácil de consensuar, de modo que se suele definir un agente por las propiedades que éste debe cumplir y que son [78]:

- **Autonomía:** Un agente puede operar sin la intervención directa del usuario y tiene algún tipo de control sobre sus acciones y su estado interno.
- **Cooperatividad:** Los agentes cooperan entre sí para lograr sus objetivos. En este caso es necesario utilizar unas reglas de comunicación estandarizadas entre agentes, definiendo un lenguaje de comunicación entre agentes (Agent Communication Language, ACL).
- **Reactividad:** El agente percibe su entorno y responde a cambios que ocurren en el mismo.
- **Proactividad:** Un agente puede tomar la iniciativa para lograr sus objetivos sin necesidad de eventos externos para reaccionar.

Otra característica que pueden presentar los agentes es la **movilidad**, de manera que podrían migrar de un sistema a otro para ejecutarse en el destino. Esta característica permitiría, entre otras cosas, reducir el ancho de banda necesario para la comunicación, disminuir la carga de los servidores en entornos cliente/servidor o extender la funcionalidad de determinados programas.

También es fundamental el concepto de negociación entre agentes, muy relacionado con la cooperatividad, que es el proceso por el que éstos llegan a un acuerdo mutuo sobre los términos y condiciones de ejecución de determinado servicio. La noción de organización virtual, como la formación dinámica de equipos de agentes que colaboran para lograr un objetivo, ha sido ampliamente estudiada en este campo y ha dado lugar al desarrollo de numeroso modelos, métodos y técnicas para establecer la cooperación

entre agentes autónomos, para asegurar la coordinación de acciones en colectivos, para seleccionar el conjunto apropiado de agentes que participen en el equipo o para especificar el modelo de confianza y seguridad en un sistema abierto.

Aunque al igual que en objetos distribuidos cada entidad tiene un estado interno determinado y la comunicación se realiza mediante el intercambio de mensajes entre ellos, hay una diferencia fundamental entre estos dos paradigmas: cuando se utilizan agentes se habla de autonomía, que implica que un agente mantiene control sobre sus acciones, mientras que una invocación a un objeto remoto no permite este grado de autonomía.

Muchos de los conceptos utilizados en el campo de los agentes para el entendimiento del entorno distribuido, como la autonomía, la negociación, el concepto de organización virtual y la cooperación, están siendo exportados a otros contextos tecnológicos.

Utilizar el paradigma de agentes puede resultar especialmente útil para la integración de sistemas heterogéneos, la gestión de recursos distribuidos y, en general, para la realización de sistemas abiertos [79]. De manera que podría ser una tecnología apropiada para la implementación de sistemas federados [80,81], aunque es necesario realizar todavía algunos esfuerzos como por ejemplo mejorar las soluciones de seguridad y privacidad de la información, especialmente cuando se utilizan agentes móviles.

En cuanto a las iniciativas para el desarrollo de esta tecnología cabe destacar la normalización de un formato para la representación de información (KIF, Knowledge Interchange Format) y de un lenguaje de comunicación entre agentes (KQML, Knowledge Query and Manipulation Language) desarrollado por la comunidad KSE (Knowledge Sharing Effort). También se ha creado un consorcio con la finalidad de promover normas para el desarrollo de aplicaciones basadas en agentes, la FIPA (Foundation for Intelligent Physical Agents, FIPA) [81].

3.2.3. Servicios Web

Al igual que las *middlewares* ya presentadas, la arquitectura de servicios Web proporciona un estándar para la interoperatividad entre aplicaciones software que pueden estar ejecutándose en distintas plataformas y entornos. En esta ocasión la interoperatividad se consigue a través de la Web, y dada la extensión de la misma esto puede facilitar la implantación de estas técnicas.

En este entorno tecnológico se entiende por servicio Web a un recurso, que se caracteriza por el conjunto abstracto de funcionalidades que proporciona, y por agente al elemento concreto de software o hardware que recibe y envía mensajes permitiendo el acceso del cliente a este recurso. Así pues un mismo servicio Web puede estar implementado por distintos agentes, pero todos ellos tendrán la misma funcionalidad sea cual sea la forma en que estén implementados o la plataforma en la que se ejecuten.

Un servicio Web se diseña para interaccionar a través de la red con otros sistemas software de forma completamente interoperable. Esta interoperatividad se consigue gracias a la normalización de tres elementos fundamentales en la arquitectura de servicios Web:

- **SOAP (Simple Object Access Protocol):** Un protocolo de comunicación basado en el intercambio de mensajes en formato XML y que se apoya en otros estándares relacionados con la Web.
- **UDDI (Universal Description, Discovery and Integration):** Un mecanismo para la publicación y localización de servicios.
- **WSDL (Web Service Description language):** Un lenguaje de descripción formal, procesable por sistemas software, que permite describir la forma de interaccionar con el agente que ofrece determinado servicio. Para ello especifica el formato de los mensajes, los tipos de datos, los protocolos de transporte utilizados, las normas para el paso de código a mensajes XML y viceversa y la localización en la que el agente puede ser invocado dentro de la red. Incluso puede dar alguna información sobre el protocolo de intercambio de mensajes que se espera.

Es importante destacar que esta descripción representa un contrato entre el agente servidor y el cliente sobre la forma de interaccionar, pero no sobre el propósito o la semántica de la información intercambiada. Esta descripción no se cubre con WSDL y no forma parte por tanto de la arquitectura básica de servicios Web. Sin embargo añadir a la descripción del servicio esta descripción de semántica permitiría una mayor automatización en la búsqueda, entendimiento y uso de los servicios, dando lugar a un sistema de servicios Web semánticos, como se presentará en el apartado dedicado a la Web Semántica.

3.2.4. Grid

La computación en Grid comienza a desarrollarse a mediados de los noventa y se fundamenta en compartir recursos para facilitar prestaciones que no serían posibles sin esta colaboración. Aunque el primer objetivo era compartir la capacidad de cómputo, permitiendo el procesado de datos a muy gran escala sobre veloces redes de área extensa de supercomputadores y clusters, actualmente se ha evolucionado hacia la idea de Organizaciones Virtuales. Se entiende la Grid como una middleware que permita compartir recursos de forma coordinada, flexible y segura entre grupos dinámicos de individuos e instituciones, a los que nos referimos como organizaciones virtuales [82,83].

La misión principal de la tecnología Grid es facilitar la coordinación para compartir recursos y la resolución dinámica de los problemas que surjan en una organización virtual que puede implicar a múltiples instituciones. Cuando se habla de compartir recursos hay que tener en cuenta que no se refiere únicamente al intercambio de archivos, sino que se considera también el acceso directo al hardware, software, datos y cualquier otro recurso que sea necesario en la resolución de un problema en colaboración. Así la Grid puede ser vista como una nueva plataforma de integración o middleware que oculta la heterogeneidad de los recursos computacionales para que éstos puedan trabajar juntos.

3.3. Descripción semántica del dominio: Ontologías

En Ingeniería del Conocimiento el término “ontología” se usa para referirse a un conjunto de conceptos, sus propiedades y las relaciones entre los mismos, que provee un entendimiento común y compartido sobre un dominio de conocimiento. Esta conceptualización, que describe la semántica del dominio, estará representada y almacenada de forma que pueda servir de soporte a diversas aplicaciones que requieren de conocimiento específico sobre la materia que la ontología representa.

La representación de la ontología debe ser pues formalizada, y estar apoyada en lenguajes normalizados que permitan la comprensión y manejo de la misma por parte de los componentes implicados en una arquitectura de comunicación para facilitar el intercambio de conocimiento y la toma de decisiones basada en éste. Esta formalización permite el desarrollo de técnicas que facilitan:

- La modificación dinámica de la ontología. Permitiendo incluir nuevos conceptos y relaciones, o eliminarlos, de modo que se puede gestionar la evolución en el tiempo de la semántica del dominio.
- El enriquecimiento de cualquier contenido incluyendo anotaciones basadas en alguna ontología.
- La transformación automática, o semiautomática, entre conceptos de distintos dominios.
- La clasificación automática de instancias y el descubrimiento automático de relaciones y equivalencias, totales o parciales, entre las mismas. Es decir permiten realizar inferencias sobre las instancias de conceptos especificados en la ontología.
- Formalizando también la representación de reglas de comportamiento de componentes se facilita el disparo de las mismas en el instante en que se cumplen determinadas condiciones en la base de conocimiento, es decir en las instancias o hechos del dominio.

A lo largo de los últimos años se han desarrollado numerosos lenguajes para la representación de ontologías, muchos de ellos muy particularizados para un dominio ya que han sido diseñados para comunidades muy específicas [84]. Sin embargo el lenguaje que está siendo aceptado como norma por la mayoría de la comunidad científica es OWL [85-89] (Ontology Web Language), desarrollado como lenguaje de propósito general y definido para ser compatible con la arquitectura de servicios Web y con la Web Semántica.

OWL se apoya en el uso de URIs (Uniform Resource Identifiers), para el nombrado y localización de conceptos (recursos), y de un lenguaje anterior, RDF (Resource Description Framework), desarrollado para la descripción de recursos y enriquecido con semántica formalizada y nuevo vocabulario para describir en más detalle los conceptos y sus relaciones y propiedades. Esto proporciona a las ontologías descritas en OWL las siguientes capacidades:

- Pueden estar distribuidas en varios sistemas.
- Tienen la escalabilidad imprescindible en las soluciones para la arquitectura Web.

- Son compatibles con otras normas Web, lo que facilita la integración.
- Son abiertas y extensibles.

En definitiva OWL facilita la interpretación del contenido publicado en la Web por agentes software, que así podrán procesar la información en lugar de simplemente presentarla a los agentes humanos, tal y como ocurre en la Web clásica.

3.4. Integración basada en la gestión semántica

3.4.1. La Web Semántica

El nivel de automatización en el descubrimiento, invocación y combinación de servicios Web es actualmente muy limitado. Las tecnologías relacionadas con la Web Semántica pretenden mejorar estas capacidades. En palabras de Tim Berners-Lee *"La Web Semántica es una extensión de la actual Web en la que la información se publica con significado formalmente definido, facilitando a computadores y personas trabajar en cooperación"* [90]. La definición formal de la semántica de la información publicada permitirá el desarrollo de nuevas aplicaciones gracias a que agentes software podrán procesar esta información de forma más eficiente, entendiendo el significado de la misma. De este modo, frente a la funcionalidad que hasta la fecha ha tenido la Web, básicamente un medio donde publicar documentos que entenderán sólo agentes humanos, la Web Semántica se convierte en el mecanismo para hacer accesible información que pueda ser procesada automáticamente por agentes software [91].

Para conseguir una interoperatividad de servicios Web a gran escala será fundamental que sus propiedades, capacidades, interfaces y efectos sean codificados de forma no ambigua e interpretable por agentes software. El marcado semántico de los servicios se apoya en el uso de ontologías que faciliten compartir, reutilizar, componer y hacer corresponder la semántica local con la de los demás. De este modo se consigue crear una base de conocimiento distribuida, donde cada agente publica su propia base de conocimiento, es decir, su visión del dominio de interés, y puede razonar sobre la de los demás servicios basándose en la semántica compartida.

Paralelamente al diseño de estas técnicas de descripción, es necesario desarrollar técnicas de agentes que exploten este marcado semántico para realizar las tareas de localización, selección, ejecución, composición y monitorización de servicios. Así, con el marcado semántico de servicios se puede especificar la información necesaria para la

localización y selección de un servicio, de modo que un agente de búsqueda pueda localizar el servicio apropiado según las necesidades de un agente cliente. Una vez localizado el servicio, su descripción puede proporcionar una API (Application Programming Interface) interpretable por software para ejecutarlo, indicando las entradas necesarias, la información devuelta y cómo ejecutar el servicio automáticamente.

Una de las grandes ventajas del marcado semántico es que se pueden localizar y utilizar servicios que sin proporcionar el objetivo buscado por el usuario puedan ayudar a conseguirlo. De este modo determinado objetivo puede provocar la creación de una serie de llamadas a agentes en cadena en las que la información se pasa de un agente a otro de manera que cada uno realiza determinados esfuerzos para conseguir el objetivo final; es decir, se facilita la composición automática de servicios para la creación de servicios más complejos.

Uno de los requisitos fundamentales para conseguir esta infraestructura es proveer de lenguajes que expresen la semántica o estructura de la información [86,92] y las reglas de inferencia para realizar razonamientos, elegir acciones o responder preguntas sobre estos datos [93]. Estos lenguajes permitirán la integración de cualquier dominio de conocimiento en el marco de la Web Semántica. Por otro lado, para conseguir funcionalidades complejas se necesita la interacción entre agentes software que colaborarán para lograr determinado objetivo, será necesario utilizar arquitecturas de comunicación que permitan esta interoperatividad. Por tanto, varias tecnologías desempeñan un importante papel en el desarrollo de la Web Semántica.

En primer lugar es necesario utilizar algún lenguaje normalizado para la representación de la ontología del dominio, es decir el lenguaje manejado por un determinado servicio. OWL (Ontology Web Language) es el elegido para su uso en la Web Semántica.

También será necesario extender la expresividad de WSDL ya que ahora es imprescindible indicar, de forma procesable por un agente software, en qué consiste exactamente el servicio. Con este fin se han desarrollado lenguajes como OWL-S [94,95] (Web Ontology Language for Services). OWL-S es una ontología que incluye los conceptos necesarios en la especificación de un servicio Web de forma suficientemente expresiva para conseguir la automatización en las tareas de descubrimiento, selección, invocación y composición de servicios [96-98]. Está especialmente indicado para la descripción de servicios Web y por tanto para aquellos

que tengan la ontología de su dominio expresada en OWL. Sin embargo podría ser utilizado para describir servicios implementados en cualquier tecnología, incluso aunque se utilizara otro lenguaje distinto de OWL para la ontología del dominio. Un agente que conozca la ontología de un dominio, y sepa interpretar OWL-S, puede saber, para los servicios en ese dominio, qué ofrecen, cómo lo hacen y cómo pueden ser invocados.

La figura 3 muestra esta ontología en el tradicional formato de tripletas (clase-relación-clase).

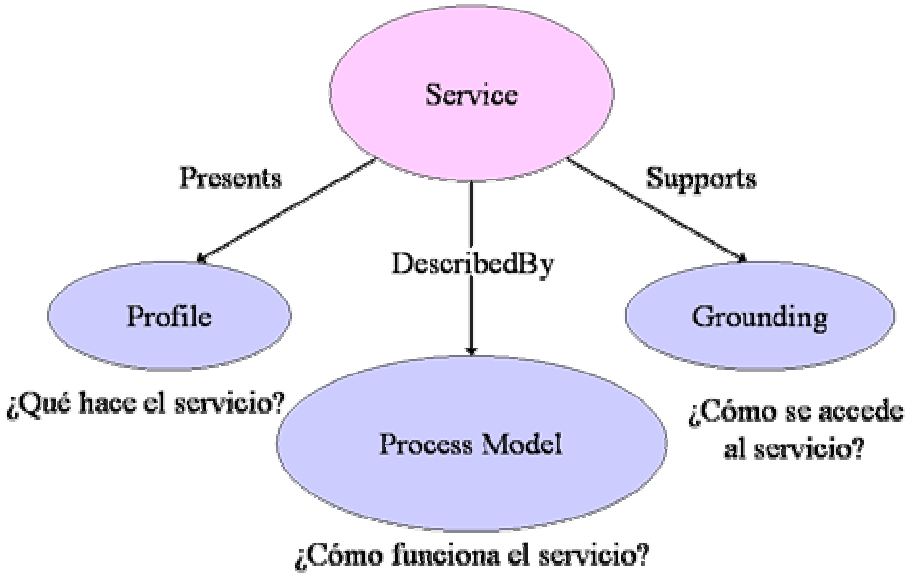


Figura 3. Modelo de servicio en OWL-S

Se puede apreciar que en OWL-S la descripción del servicio se hace desde tres puntos de vista:

- **El perfil o *Profile*:** que muestra la funcionalidad y características del servicio, indicando qué ofrece al cliente y qué requisitos exige a éste.
- **El modelo de proceso o *Process model*:** que describe la actividad del proveedor del servicio y cómo éste ofrece las prestaciones indicadas al cliente, es decir qué ocurre cuando el servicio es invocado.
- **El modelo de comunicación o *Grounding*:** que indica cómo el agente cliente puede interactuar con el servidor. Normalmente especifica un protocolo de comunicación, el formato de los mensajes intercambiados y otros detalles como

el número del puerto por el que se contacta con el agente servidor. Además para cada tipo abstracto especificado en el modelo de proceso se describe el procedimiento de transformación de código a mensajes XML, y viceversa, detallando el formato necesario para el intercambio de datos de ese tipo.

En cuanto a la infraestructura de comunicación, la arquitectura de Servicios Web será el estándar para la interoperatividad entre agentes software, que pueden estar ejecutándose en distintas plataformas. Se publican las tareas o esfuerzos realizados por cada agente como servicios y se especifica el modo de acceder a éstos, pero no cómo se implementan y tampoco se imponen restricciones en cómo combinar distintos servicios Web. De manera que aún quedan esfuerzos por realizar, por ejemplo cómo publicar la descripción en OWL-S en el UDDI para facilitar las tareas de descubrimiento y selección de servicios [99,100], qué lenguajes pueden soportar la composición automática de servicios [97,101], cómo afrontar las tareas de seguridad en un entorno tan complejo [102] o qué lenguajes pueden utilizarse para la especificación de reglas de comportamiento de los agentes [93].

Así, la fusión de las tecnologías de gestión de ontologías, más concretamente aquellas expresadas en OWL, y de Servicios Web converge hacia el desarrollo de la Web Semántica, donde se pueden desarrollar servicios cuya funcionalidad está expresada en términos de ontologías, lo que la hace "comprensible" a un agente software cliente que podrá tomar decisiones de actuación para lograr los objetivos buscados.

3.4.2. Aportaciones y limitaciones de la Web Semántica a la integración de sistemas de información sanitarios.

El uso de las tecnologías presentadas en este capítulo puede suponer un gran avance en la integración de sistemas de información de cualquier naturaleza y por supuesto de los sanitarios. De hecho existen varios trabajos que tratan de aplicar estas técnicas en el campo de la informática sanitaria [103-107]. Aunque desde hace tiempo se vienen desarrollando tecnologías *middleware* que facilitan la interconexión, el avance en la integración de sistemas de información desarrollados independientemente se ha visto fuertemente limitado por la incompatibilidad semántica de los mismos. La resolución de estos conflictos sólo puede solucionarse unificando los conceptos manejados en los distintos dominios, lo que hasta ahora solía obligar a modificar el funcionamiento de los

sistemas a integrar e incluso el esquema de base de datos subyacente. Esto supone un costoso esfuerzo en muchas ocasiones imposible de asumir.

El modelo deseado es una arquitectura federada en la que los sistemas sigan funcionando como antes de la integración y en la que se simplifiquen los procedimientos de integración. De manera que aunque hace años que existe tecnología que permite la interconexión de sistemas, una integración rentable de sistemas de información heterogéneos necesita todavía que maduren las técnicas de integración semántica. La mayor evolución de éstas se está dando en el campo de la Web Semántica, y como los servicios Web semánticos combinan de forma muy adecuada tecnologías de comunicación de agentes software y de gestión de conocimiento se sitúan como fuertes candidatos a ser la tecnología seleccionada en una arquitectura de sistemas sanitarios.

Las funciones y modo de acceso a los servicios disponibles en la organización deberían ser descritos y publicados para que puedan ser utilizados por agentes software autorizados. Esta descripción debe estar basada en términos acordados por la comunidad en la que se pretenden integrar los sistemas, una ontología compartida. De este modo el esfuerzo de integración es la descripción de los servicios ofrecidos conforme a la ontología del dominio compartido y el desarrollo de agentes que permitan el acceso basado en los protocolos de servicios Web.

De manera que a la hora de seleccionar la base tecnológica de una organización sanitaria que pretenda integrar todos sus sistemas de información, esta tecnología debe ser considerada especialmente por dos hechos: primero porque los mayores avances en gestión semántica se están dando en este campo, sobre todo por el gran número de investigadores implicados, de modo que estas innovaciones podrán ser reutilizadas en el escenario de interés; y segundo porque muchos trabajos de investigación actuales, tanto en el terreno de la informática clínica como en la bioinformática, se están desarrollando utilizando algunas de las tecnologías relacionadas con la Web Semántica [12,107-113] y por tanto la explotación de los resultados obtenidos pasará por el manejo de las mismas.

Sin embargo es necesario considerar que las prestaciones necesarias en una organización sanitaria son especialmente complejas y que la plataforma de Web Semántica puede no cubrir todas las necesidades de la misma.

En especial es de destacar que en las tecnologías de servicios Web semánticos las relaciones entre agentes están limitadas al paradigma cliente/servidor, mientras que cada vez es más habitual que dentro de la organización se integren aplicaciones que requieren de la gestión de flujos multimedia, con requisitos específicos de calidad de servicio que la plataforma debe asegurar. El esfuerzo para incluir estas prestaciones sobre una arquitectura basada en Web Semántica puede ser considerable.

Por otro lado, la complejidad de las tareas a realizar puede necesitar de la consideración de estados globales del sistema, que condicionen el comportamiento de algunos de los componentes de la arquitectura. Sin embargo la Web Semántica no está pensada, por ahora, para gestionar estados globales.

3.4.3. La Grid Semántica

El concepto de Grid Semántica se introdujo en 2001 [74] y se entiende como una extensión de la actual Grid en la que el significado de la información y los servicios se expresan formalmente facilitando a máquinas y personas trabajar en cooperación. Surge de la intersección de investigaciones en los campos de la Grid, Web Semántica y comunidades de agentes software, para tratar de eliminar el escalón entre teoría y práctica que existe en la computación en Grid [83,114].

De la Web Semántica se hereda la visión de facilitar que los datos expuestos en la Web sean entendibles por las máquinas, de manera que las tecnologías de gestión de información y conocimiento desarrolladas en este campo podrían ser aplicadas también a la Grid. Por otro lado, los avances en computación basada en agentes permiten evolucionar de una filosofía de Grid fundamentada en una arquitectura de servicios a una computación soportada por componentes software autónomos que pueden actuar e interaccionar de forma flexible. De manera que la grid semántica podría ser entendida como un puente para unir las comunidades centradas en la gestión de conocimiento y la computación basada en agentes, con las emergentes que desarrollan las tecnologías de Web Semántica y computación ubicua [115].

Algunos de los requisitos fundamentales de esta tecnología son:

- Descripción, descubrimiento y uso de recursos.
- Descripción y ejecución de procesos.
- Autonomía de comportamiento.

- Mecanismos de gestión de seguridad y confianza.
- Gestión de anotaciones.
- Integración de información.
- Facilidades para gestión de flujos y vistas de información.
- Gestión del conocimiento basada en el contexto.
- Gestión de comunidades.
- Incluir prestaciones para entornos inteligentes y computación ubicua.
- Facilidad de configuración y desarrollo.
- Facilitar la integración de sistemas ya desarrollados.

La nueva arquitectura de Grid Semántica extiende la Grid tradicional con nuevos componentes para la gestión de conocimiento [115], como se muestra en la figura 4, donde se distingue entre una "Grid del conocimiento", con los elementos para gestionar el conocimiento generado por aplicaciones que utilizan y explotan la Grid, y una "Grid semántica", que contiene los elementos necesarios para realizar la gestión y ejecución de los componentes de la arquitectura.

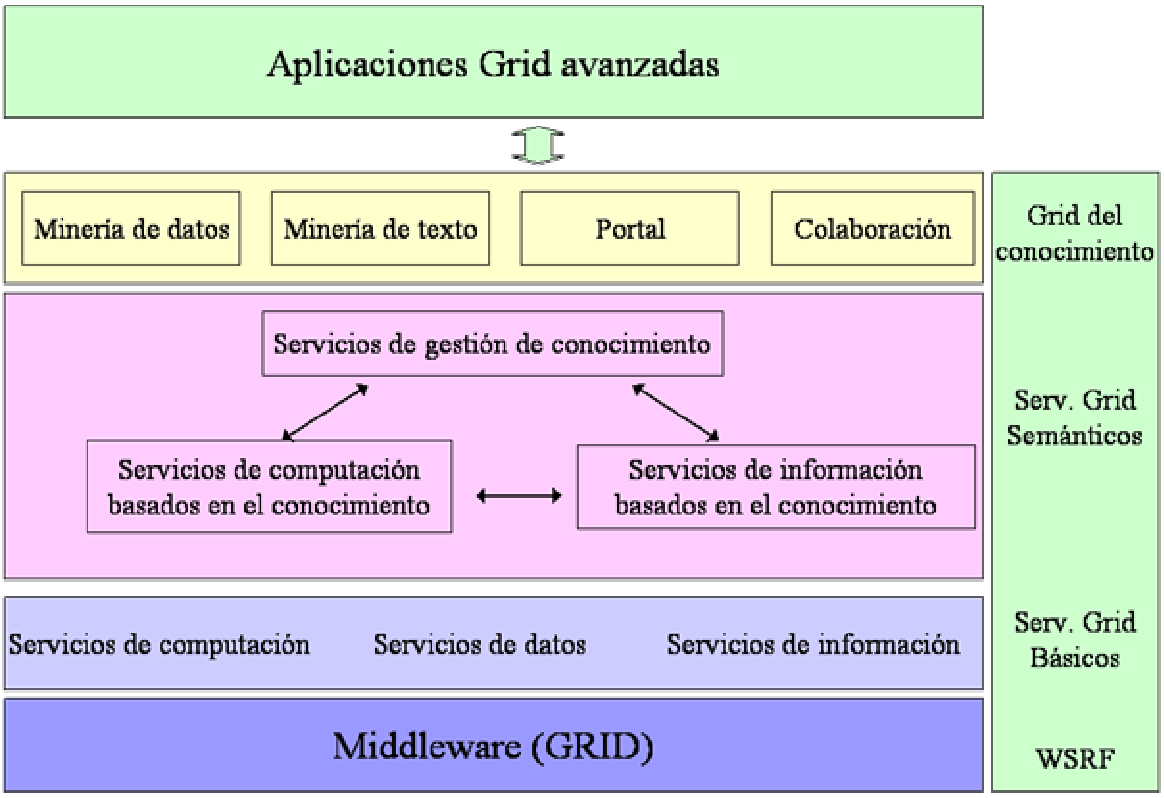


Figura 4. Arquitectura de la GRID Semántica

3.4.4. Aportaciones y limitaciones de la Grid Semántica a la integración de sistemas de información sanitarios.

La visión de la Grid semántica extiende la funcionalidad de la Web Semántica en los dos aspectos que se habían señalado como posibles problemas al seleccionar esta tecnología para el soporte de la arquitectura sanitaria: la gestión de estados globales y de paradigmas distintos al tradicional cliente/servidor. Sin embargo, la investigación en este campo está más retrasada y hay menos grupos implicados que en la Web Semántica, por lo que es difícil considerarla a corto plazo, pero sí conviene tenerla muy en cuenta para las tareas de investigación más centradas en un medio y largo plazo. Es de destacar que los más prestigiosos investigadores en estas tecnologías las aplican precisamente al dominio sanitario o a la bioinformática [111,116,117].

3.5. Conclusiones

La interoperatividad de sistemas heterogéneos y el desarrollo de entornos federados deberá estar soportado por alguna tecnología que facilite la computación distribuida. Estas tecnologías se podrían definir de forma genérica como el conjunto de técnicas que permiten la gestión de componentes software y de los mecanismos de interacción entre los mismos con la finalidad de ofrecer facilidades a los usuarios o a otros componentes software. La flexibilidad y escalabilidad necesarias para soportar la evolución de la arquitectura facilitando la integración de nuevos miembros a la federación, y el desarrollo de nuevas aplicaciones sobre la misma, dependerá de la tecnología utilizada y se incrementará cuanta mayor capacidad de gestión semántica se incorpore.

En cuanto a la elaboración de métodos y modelos abiertos para la federación de sistemas sanitarios, la vigencia de los mismos de forma transparente a fronteras técnicas y/o temporales sólo estará asegurada si estos son independientes de la tecnología que se elija para su implementación física.

Capítulo 4. Métodos y modelos

4.1. Introducción

En este capítulo se analizan los principales referentes sobre el modelado de sistemas de información y de sistemas distribuidos utilizados en esta Tesis Doctoral:

- **El modelo dual**, fundamento del concepto de arquetipo, que es la base de las soluciones propuestas por los organismos de normalización para solventar los problemas de incompatibilidad semántica en el entorno sanitario [25-27,71].
- **El modelo ODP** para la normalización de sistemas de procesamiento distribuido abiertos [15,118-128].
- **La arquitectura HISA**. Una particularización del modelo ODP especificada por el CEN. Dada la naturaleza heterogénea y distribuida de los sistemas de información clínicos, el CEN impone la utilización de ODP para la descripción de arquitecturas que busquen la interoperatividad y federación de sistemas [8,44,45].

4.2. El modelo dual

4.2.1. Introducción

El propósito de todo sistema de información es **la creación y procesado de instancias de conceptos.** Los **conceptos** son tipos, o clases, de entidad manejados por el sistema. Las **instancias** son casos reales y concretos de los conceptos, normalmente en forma de datos estructurados representando una entidad en particular, es decir, son objetos.

Los sistemas de información clásicos se han desarrollado de tal forma que los conceptos que los sistemas tienen que gestionar son directamente codificados en su software y en sus bases de datos. Esta forma permite un desarrollo relativamente rápido, pero su vida es limitada y son difíciles de modificar y ampliar. Incluso en sistemas orientados a objetos, que permiten un modelado muy eficiente, existe un problema básico: el software nunca se puede dar por finalizado ya que continuamente cambian o aparecen

nuevos conceptos en el dominio, obligando continuamente a la reconstrucción, prueba y ampliación de sistemas que ya están en funcionamiento [25,129].

Se enumeran a continuación una serie de limitaciones que suelen tener los sistemas diseñados mediante el modelado clásico:

- El modelo sólo codifica los requisitos encontrados durante la etapa de desarrollo, junto con las mejores suposiciones de los requisitos futuros.
- Suelen contener conceptos genéricos y específicos del campo de aplicación del sistema (dominio) en el mismo nivel jerárquico, y eso puede llegar a ser problemático.
- A menudo es difícil completar los modelos satisfactoriamente, ya que el número de conceptos de dominio puede ser muy elevado y el sistema suele ser difícil de extender a posteriori.
- Puede haber problemas con la implementación de la semántica. A menudo no es posible modelar de forma clara los conceptos del dominio directamente en clases, métodos y atributos. Además éstos tienen una variabilidad significativa, y es usual que requieran restricciones expresadas en forma de predicados lógicos para completar su definición.
- El modelado puede ser difícil debido al hecho de que hay dos tipos de actores involucrados: especialistas en los conceptos de interés en el sistema y los desarrolladores del mismo. Ambos tienen que trabajar con métodos y lenguajes con los que no están familiarizados, y establecer mecanismos de colaboración adecuados no suele resultar sencillo.
- La introducción de nuevos conceptos requiere cambios en el software, reconstrucción, depuración y nuevo despliegue, lo cual es caro y arriesgado.
- La interoperatividad es difícil de conseguir, ya que cada sistema debe continuamente reconstruir sus modelos para ser compatible con los otros, o modernizar sus interceptores, que son los elementos software que le permiten integrarse en una arquitectura de sistemas.
- La estandarización es difícil de lograr. Con modelos de dominios grandes es, desde el punto de vista lógico y tecnológico, difícil para diferentes proveedores y usuarios ponerse de acuerdo sobre un modelo común. La falta de

estandarización no sólo dificulta la interoperatividad, sino que hace el procesado automático casi imposible ya que no hay ninguna suposición general que los sistemas puedan hacer sobre el modelo subyacente.

La razón por la que el modelado clásico no resulta satisfactorio es porque el diseño de los sistemas de información se plantea para cubrir las necesidades específicas de los usuarios a corto o medio plazo sin considerar el crecimiento a largo plazo ni la interoperatividad del sistema. Así, tanto los esquemas de las bases de datos como el software desarrollado, manejan conceptos de alto nivel, es decir muy próximos al usuario, que pueden llegar a ser muy variables en el tiempo.

En otras palabras, tanto el mantenimiento y ampliación de los sistemas como la validación de la información son directamente dependientes de las entidades software utilizadas en el desarrollo del sistema. Como consecuencia de este paradigma de diseño la mayoría de los sistemas de información no sólo no sirven bien a sus usuarios a largo plazo, sino que además muestran una interoperatividad limitada.

Se necesita un planteamiento diferente que se ajuste a la idea de que el mundo es un lugar que cambia, dinámico. En muchos dominios tanto el número de conceptos como la tasa de cambios de los mismos son elevados, y el caso de la sanidad es un ejemplo claro. El número de conceptos implicados en el dominio sanitario es muy amplio. El gran esfuerzo realizado en el desarrollo de tesauros [61,63,64] e integración de los mismos [130-132] y la continua evolución de estos sistemas clásicos de terminología clínica así lo evidencian. La tasa de cambios es difícil de medir, pero hay al menos cinco factores que contribuyen a dichos cambios:

- Muchos conceptos no son modelados de una forma correcta debido a que no se dispone de un método o lenguaje apropiado.
- Surgen nuevos conceptos y otros cambian debido a la mejora de la tecnología y la técnica. Así, por ejemplo, surgen nuevos tipos de dispositivos que permiten realizar medidas nuevas y más complejas.

- Nuevas ciencias, como es el caso de la virología, la genética o la proteómica[1], abren nuevos caminos creando dominios de conceptos en el ámbito sanitario y que hay que considerar en los sistemas de esta índole.
- Cambios en el manejo de datos operacionales. Especialmente en sanidad los cambios en los procedimientos a seguir implican modificaciones en los modelos de información clínicos.
- Cambios en los procesos de negocio, de información y de seguridad debido a la legislación que también implican modificaciones en los modelos de información.

La aproximación propuesta por el openEHR [26,71] (y su predecesor GEHR [25]) es un riguroso modelado del conocimiento basado en un principio básico: **la separación en los sistemas de información de los asuntos relacionados con la semántica del dominio de aquellos relacionados con los aspectos técnicos.** Esto se traduce en:

- La eliminación de los conceptos del dominio, cercanos al profesional sanitario, del software y de los modelos de base de datos para llevar un control independiente de dichos conceptos y así poder usar un vocabulario estandarizado y librerías de modelos de conceptos relacionados con la sanidad.
- Reestructurar el software y las bases de datos usando una arquitectura de sistema basada en modelos de referencia de objetos, y diseñada para procesar la información usando unas definiciones de dominio suministradas externamente.

En este nuevo paradigma de desarrollo de sistemas de información, el análisis y el diseño no expresan semántica de alto nivel, por ejemplo de términos clínicos en el caso de sistemas sanitarios, sino únicamente conceptos de estructuración de la información con los cuales se puede expresar cualquier concepto del dominio. Estos conceptos estructurales se especifican en el modelo de información, también denominado modelo de referencia, de manera que cualquier entidad gestionada en el sistema podrá identificarse como una instancia de alguno de estos conceptos. Esto incluye modelos como el dado en la norma 13606 del CEN [39], el RIM del HL7 [47] o los modelos de información del openEHR [33,66,68-70].

---

[1] Dada la reciente aparición de este término, 1990, aún no está incluido en las correcciones a la XXIII edición del Diccionario de la Lengua Española de la RAE

4.2.2. Arquetipos

Como se ha dicho anteriormente, el propósito fundamental de los sistemas de información es gestionar instancias de conceptos. La metodología aquí presentada, orientada a la gestión del conocimiento, propone que la definición de conceptos del dominio sea separada del modelo técnico. La dificultad está en encontrar un mecanismo para formalizar la gran complejidad de los conceptos del mundo real y permitir así su gestión tanto a humanos como a agentes software. Con este fin se desarrolla la idea de arquetipo.

En esta "reconcepción" de sistemas de información, los conceptos del dominio son modelados formalmente como arquetipos, usados por el sistema en tiempo de ejecución y aportan la flexibilidad necesaria para los conceptos más dinámicos y más cercanos al personal sanitario. Por otro lado el software será una implementación del modelo de información o de referencia, que proporciona un marco muy general para los conceptos más rígidos del dominio clínico y más próximos al ingeniero de sistemas. Un arquetipo es una expresión computable de un concepto del dominio, que puede asimismo tener una gran complejidad interior, representada en forma de declaraciones de restricciones estructuradas, basadas en algún modelo de referencia.

Los propósitos fundamentales para la creación de arquetipos son:

- **Comunicación humana:** permiten que los conceptos de dominio sean modelados de una manera formal por los expertos de dominio, evitando que estos tengan que entrar en la complejidad técnica subyacente.
- **Vistas especializadas:** permiten la personalización del conocimiento gestionado. Al facilitar que la información sea manejada de manera conforme a arquetipos especializados se tiene una visión particular de la misma.

Los beneficios claves de los arquetipos incluyen:

- **Sistemas con capacidad de gestión de conocimiento:** permiten a los sistemas de información guiar y validar la información introducida por un usuario en tiempo de ejecución, garantizando que todas las instancias de información (datos con un valor concreto) creadas están conformes a los requisitos del dominio. La separación entre los asuntos de información y los de conocimiento en los sistemas software permite reducir el coste del software y hacerlo más duradero.

- **Interoperatividad a nivel de conocimiento:** la capacidad de los sistemas para comunicarse de una manera fiable con otros sistemas a un nivel de conceptos de conocimiento y no únicamente a nivel de estructuras.
- **Potenciación de los dominios:** la potenciación de los especialistas de dominio para definir los conceptos con los que ellos trabajan y tener control directo sobre sus sistemas de información.
- **Búsquedas inteligentes:** son usadas en tiempo de ejecución para permitir búsquedas eficientes de datos basadas en la estructura del arquetipo adecuado.

La idea subyacente es que un concepto puede verse como una definición de restricciones estructurales; es decir, que un conjunto de restricciones puede servir para definir las instancias que se corresponden a un determinado concepto del dominio. A este conjunto de restricciones se le llama arquetipo porque puede considerarse como un modelo prototipo que contempla numerosas variaciones posibles. Así un arquetipo es una constelación de combinaciones válidas de objetos del **modelo de información** para un concepto de dominio particular.

Esta nueva concepción de los sistemas de información basada en arquetipos es clave para lograr una amplia difusión e interoperatividad a nivel de conocimiento. Garantiza que toda la información creada está conforme a modelos del dominio, descritos en los arquetipos, asegurando así que todos los sistemas cooperantes pueden entender la información al más alto nivel posible. En particular, esto permite un procesado automático efectivo, ya que los sistemas podrían hacer suposiciones sobre la estructura y el contenido de la información inspeccionando los arquetipos usados para crearla.

Esta idea ha sido adoptada por el CEN [27], que colabora estrechamente con grupos como el openEHR, muy implicados en el desarrollo de modelos que ayuden a la interoperatividad de sistemas, y es base del proyecto de norma europea 13606.

Se pueden definir arquetipos en distintos niveles ontológicos: la composición permite crear arquetipos de nivel superior a partir de otros de nivel inferior. Por otro lado los conceptos están también relacionados con otros términos de especialización. La especialización puede ocurrir debido a alteraciones localizadas en conceptos estándares al particularizar en determinado contexto. Por ejemplo, unos exámenes médicos generales son más básicos o menos específicos que realizados por un especialista de medicina interna.

Los arquetipos pueden ser usados directamente para los propósitos computacionales, para los cuales suelen ir encapsulados en plantillas.

Una plantilla es un mecanismo usado localmente en los sistemas para la validación o creación de datos. Semánticamente las plantillas son restricciones sobre los arquetipos y tienen una relación 1:N con los conceptos subyacentes, cada uno de los cuales está descrito con un arquetipo.

Las plantillas tienen dos propósitos principales:

- **Construcción de datos:** son usadas en tiempo de ejecución para restringir la creación de datos en contextos locales para ajustarse a los requisitos de la captura de datos.
- **Validación de datos:** son usadas para validar datos de otras fuentes en tiempo de ejecución.

Mientras que los arquetipos son generalmente modelos tolerantes y tienen posibilidades de composición muy abiertas, las plantillas son usadas para limitar las elecciones de los arquetipos para propósitos locales o específicos. Las plantillas pueden ser usadas para controlar los siguientes aspectos:

- Composición de arquetipos.
- Reducción de los términos permitidos.
- Restricción de la opcionalidad.
- Borrado de estructuras definidas en los arquetipos referenciados.
- La presentación visual.

La figura 5 resume estas ideas y en ella se pueden apreciar las siguientes relaciones:

- Los datos son instancias de un modelo de información. Es decir, que los datos que el sistema maneje deberán ser objetos conformes al modelo de clases diseñado para el mismo.
- Los arquetipos son instancias de un modelo de arquetipos, el cual es un formalismo común para expresar todos los arquetipos. Esto significa que todos deberán cumplir las normas impuestas por el modelo de arquetipos vigente en el sistema.

- El modelo de arquetipos está formalmente relacionado con el modelo de información, de tal manera que su semántica son restricciones sobre objetos de tipos definidos en el modelo de información. También puede incluir elementos lingüísticos permitiendo que sean expresadas relaciones entre elementos e invariantes.
- Si los datos son creados y modificados usando arquetipos, estos restringen la configuración de las instancias de datos para que sean válidas según el arquetipo. Así los datos cumplirán siempre las restricciones impuestas en el arquetipo, lo que puede permitir por un lado la identificación y resolución de conflictos en la asignación de valores y por otro puede ayudar a la identificación de la semántica de determinada "pieza" de información facilitando así la interoperatividad entre sistemas heterogéneos.

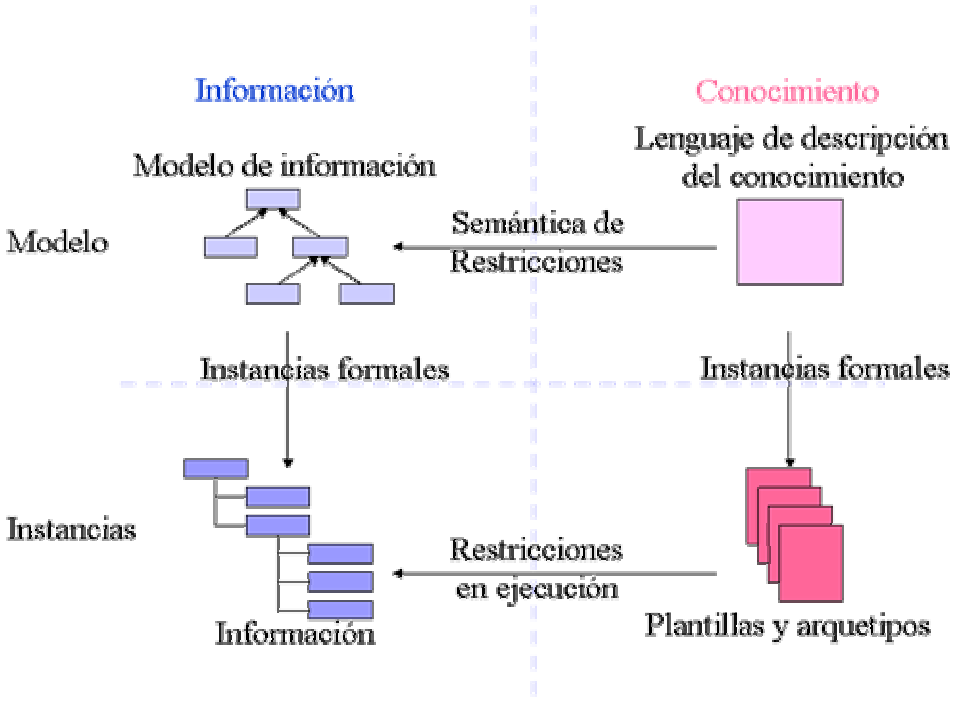


Figura 5. Arquitectura del modelo de arquetipos

4.2.3. Conclusiones

En esta Tesis Doctoral no se diseñará un sistema de información, de manera que no se va a utilizar estrictamente el modelo dual, pero sí la idea del uso de arquetipos para la expresión de semántica de alto nivel, y de modelo de información o de referencia para fijar las estructuras comunes y rígidas del dominio clínico. De manera que la vista de información de la arquitectura será diseñada en dos capas en lugar de en una sola, como

tradicionalmente se viene haciendo. Además se incorporarán a la arquitectura componentes para la gestión de la segunda capa del modelo de información, es decir de los arquetipos.

Para la descripción de estos arquetipos openEHR ha desarrollado un lenguaje de propósito muy específico: ADL (Archetype Definition Language). Esto es una gran aportación dado que al estar diseñado a medida para la especificación de arquetipos cubre todas las necesidades identificadas. Sin embargo, existen lenguajes para la descripción de la semántica de un dominio (ontología) que son de propósito más general y que están normalizados por organismos mundialmente aceptados por la comunidad científica en materia de tecnologías de la información y comunicaciones. Es el caso, por ejemplo, de OWL (Ontology Web Language). El uso de este lenguaje para la descripción de arquetipos podría tener, entre otras, dos ventajas importantes:

- Aprovechar los avances técnicos que se realicen para otros sistemas de información, no clínicos, especialmente las técnicas de gestión semántica que actualmente evolucionan muy rápidamente en el campo de los servicios Web, donde se utiliza OWL.
- Facilitar la interoperatividad de los sistemas de información clínicos con sistemas de información de diferentes dominios pero íntimamente ligados a ellos, como los pertenecientes a compañías aseguradoras o de gestión de personal y recursos.

Dado que se buscan precisamente soluciones de integración abiertas, durante la elaboración de esta Tesis la capa de conocimiento, es decir los conceptos de alto nivel, se han expresado en OWL en lugar de en ADL, lo que ha servido para identificar las limitaciones de este lenguaje y las posibles soluciones a las mismas.

4.3. Open Distributed Processing

4.3.1. Introducción

ODP es un conjunto de recomendaciones de la serie X (Redes de datos y comunicación entre sistemas abiertos) de la ITU-T [15,118-128] cuyo propósito es la elaboración de normas que permitan beneficiarse de la distribución de los procesos de gestión de la información dentro de entornos heterogéneos y a través de múltiples dominios pertenecientes a distintas organizaciones. Estas normas imponen restricciones a la

especificación de sistemas y proveen de una infraestructura que tiene en cuenta las dificultades inherentes al diseño y la programación de sistemas distribuidos.

Los sistemas distribuidos son importantes porque existe una necesidad creciente de interconectar sistemas de procesamiento de la información que surge como consecuencia de tendencias organizacionales que exigen un intercambio de información tanto entre grupos en el interior de una organización como entre organizaciones cooperantes. La construcción de tales sistemas no es fácil. Requiere una arquitectura y, dado que una solución técnica única no satisfará todas las exigencias, dicha arquitectura deberá ser flexible. Además, puesto que un solo proveedor no tendrá respuesta para todas las cuestiones, es esencial que la arquitectura y toda función necesaria para implementarla estén definidas en un conjunto de normas, de modo que varios proveedores puedan colaborar en la provisión de sistemas distribuidos.

ODP provee de una infraestructura de componentes y funciones que permiten solventar algunas dificultadas inherentes al diseño y la programación de sistemas distribuidos como son:

- **Interacción remota:** los componentes de un sistema distribuido pueden estar separados en el espacio de manera que las interacciones entre ellos pueden ser locales pero también remotas.
- **Concurrencia:** cualquier componente de un sistema distribuido puede ejecutarse en paralelo con cualquiera de los demás.
- **Carencia de estado global:** el estado global de un sistema distribuido no puede ser determinado con precisión.
- **Fallos parciales:** cualquier componente de un sistema distribuido puede fallar independientemente de los demás.
- **Asincronía:** las actividades de comunicación y procesado no son gobernadas por un reloj global. No puede asumirse que los cambios en el sistema puedan tener lugar en un instante determinado.

Como en ODP se considera que los componentes del sistema pueden estar distribuidos en varios dominios pertenecientes a distintas organizaciones y que utilizan distintas tecnologías, existe una clara descentralización, careciéndose de un punto central de control. Esto añade a dichos sistemas algunas características muy especiales:

- **Heterogeneidad:** en un sistema abierto y distribuido no hay garantía de que los componentes estén basados en la misma tecnología y además las distintas tecnologías irán cambiando a lo largo del tiempo. Esta heterogeneidad se puede dar en cualquier nivel: hardware, sistemas operativos, redes y protocolos de comunicación, lenguajes de programación, funcionalidad, etc.
- **Autonomía:** un sistema abierto distribuido puede incluir varias autoridades autónomas de control o gestión, no existiendo un control único. Esta autonomía viene impuesta porque los recursos de procesado están bajo el control de distintas organizaciones.
- **Evolución:** durante el tiempo de vida de un sistema distribuido hay que asumir un gran número de cambios motivados por el progreso tecnológico, decisiones estratégicas surgidas de nuevos objetivos y nuevas funcionalidades que se incluyan en el sistema.
- **Movilidad:** las fuentes de información, los nodos de procesado y los usuarios podrían tener movilidad física. Los datos y el software podrían también ser intercambiados entre nodos para conseguir esta movilidad física o para mejorar la eficiencia del sistema.

Con la finalidad de incluir todas estas características, los sistemas diseñados según las recomendaciones de ODP tienen las siguientes propiedades:

- **Sistemas abiertos:** en el sentido de que permiten, por un lado, la portabilidad, haciendo posible la ejecución de los componentes en distintos nodos de procesado sin necesidad de hacer modificaciones; y, por otro lado, la interoperatividad, permitiendo soportar interacciones bien definidas entre componentes que pueden residir en distintos sistemas.
- **Integración:** permitiendo incorporar varios sistemas y recursos a uno solo sin suponer un coste elevado a los desarrolladores. Así se considera, por ejemplo, la integración de sistemas con distintas arquitecturas, recursos o comportamientos. Conseguir esta propiedad es complicado debido a la heterogeneidad.
- **Flexibilidad:** permitiendo la evolución de los sistemas. Un sistema abierto distribuido debe ser capaz de permitir, por ejemplo, cambios en tiempo de

ejecución, y podrá ser reconfigurado dinámicamente para acomodarse a cambios y circunstancias.

- **Modularidad:** es la base de la flexibilidad, se considerará que las partes del sistema son autónomas pero están interrelacionadas.
- **Federación:** permitiendo combinar sistemas que se encuentran en distintos dominios técnicos o administrativos para lograr un objetivo común.
- **Gestionabilidad:** permitiendo monitorizar, controlar y gestionar los recursos de un sistema para poder facilitar las tareas de configuración, gestión de calidad de servicio y las políticas de tarificación.
- **Provisión de calidad de servicio:** pudiéndose establecer un conjunto de requisitos de calidad en el comportamiento del sistema. Así por ejemplo se pueden especificar requisitos de tiempo o disponibilidad en el contexto de recursos remotos e interacciones. Además se pueden incorporar mecanismos de tolerancia a fallos que permitan al sistema continuar funcionando si alguno de los componentes falla. Estos mecanismos son imprescindibles en grandes sistemas distribuidos donde es difícil que todos los componentes del sistema estén siempre operativos simultáneamente.
- **Seguridad:** las facilidades e información de los sistemas están protegidos contra accesos no autorizados. Los requisitos de seguridad son difíciles de conseguir a causa de las numerosas interacciones remotas y de la capacidad de movilidad, de los componentes o usuarios, que puede haber en estos sistemas.
- **Transparencia:** los detalles y mecanismos utilizados para superar todos estos problemas inherentes a la distribución deben ser transparentes a las aplicaciones. Este requisito es fundamental en la construcción de aplicaciones distribuidas. Todos los aspectos de distribución deben ser enmascarados, total o al menos parcialmente.

4.3.2. El modelo de referencia ODP (RM-ODP)

Se necesita un marco de referencia para la especificación del sistema distribuido que facilite la estructuración de las características del mismo en servicios comunes. La especificación de un servicio debe recoger la forma de representar datos, los protocolos de transporte, los algoritmos de localización de proveedores de servicios, etc. Es difícil

conseguir una infraestructura común que incluya todas las propiedades requeridas en un sistema, sobre todo cuando el tamaño del mismo es considerable. Así que es necesario seleccionar los componentes que cubran los requisitos de tipos particulares de aplicaciones y especificar un marco de referencia que describa los componentes de esta infraestructura y muestre cómo se relacionan.

De este modo el objetivo de la estandarización basada en ODP es el desarrollo de un marco de referencia para la especificación de sistemas y de la infraestructura de componentes correspondiente.

El modelo de referencia (RM-ODP) proporciona este marco, permitiendo el desarrollo de estándares basados en ODP que especifican componentes comunes consistentes y que pueden ser combinados para conseguir infraestructuras que proporcionen los requisitos de un usuario. Para ello se requiere del esfuerzo y la coordinación de los desarrolladores de sistemas y estándares. Si se siguen los principios de la arquitectura y el estándar ODP en la construcción de sistemas distribuidos el resultado será que además los sistemas serán abiertos.

El modelo de referencia define:

- La división de la especificación del sistema en vistas, que simplificarán la descripción de sistemas complejos.
- Un conjunto de conceptos generales para representar estas vistas.
- Un modelo de infraestructura que soporta estos conceptos, proporcionando transparencia de distribución, ofertados como herramientas de especificación.
- Los principios para asegurar la conformidad de sistemas basados en ODP.

Los grupos Open Software Foundation (OSF) y Object Management Group (OMG) ya habían publicado arquitecturas que soportaban la combinación de funciones para proporcionar sistemas distribuidos. El modelo de referencia de ODP tiene el valor añadido de considerar la federación, transparencia y gestión del sistema y de definir un modelo, basado en puntos de referencia, que facilita la integración de funciones desde distintas fuentes. Así que el modelo de referencia de ODP permite que sistemas distribuidos con tecnologías dispares sean integrados con soluciones técnicas de bajo coste para cumplir los requisitos del mercado.

4.3.3. Vistas

La especificación completa de cualquier sistema distribuido no trivial implica una gran cantidad de información. Intentar incluir todos los aspectos del diseño en una única descripción es, generalmente, imposible. La mayoría de las metodologías de diseño intentan establecer un conjunto de modelos de manera que cada uno represente una faceta del diseño y que satisfaga los requisitos concernientes a alguno de los grupos implicados en el proceso de modelado.

En ODP esta separación se hace identificando 5 vistas y asociando a cada una un lenguaje que exprese los conceptos y reglas relativos al sistema que pueden ser descritos desde esa vista. El conjunto de vistas se ha elegido para que sea a la vez simple y completo y para que cubra todos los aspectos necesarios en el diseño de la arquitectura. A continuación se presentan estas vistas.

Vista de empresa

Relativa a las funcionalidades de negocio. Modela el sistema y el entorno en el que interacciona describiendo el papel del mismo dentro de la infraestructura de determinado negocio, el papel de los clientes respecto al sistema y las políticas de negocio.

Vista de información

Relativa a la información que es necesario gestionar en el sistema. Proporciona una vista común y consistente de la información que contiene un sistema y del procesado que lleva a cabo con la misma. Puede ser referenciado para especificar los componentes del sistema y el flujo de información entre ellos.

Vista computacional

Relativa a la descripción del sistema como un conjunto de objetos que interactúan a través de sus interfaces, permitiendo así la distribución. Modela el sistema en términos de los componentes lógicos que lo conforman y que serán fuentes y/o sumideros de información. El uso de lenguajes y especificaciones de computación puede cubrir los requisitos de un amplio espectro de sistemas distribuidos, proporcionando el máximo potencial de portabilidad e interfuncionamiento y permitiendo la definición de constantes de distribución sin especificar los detalles de los mecanismos implicados.

Vista de ingeniería

Relativa a los mecanismos que soportan la distribución. Define la infraestructura de computación en red que proporciona la estructura definida en el modelo computacional y la transparencia de distribución necesitada.

Vista de tecnología

Relativa a los detalles de los componentes que forman el sistema; es decir, cómo está estructurado en términos de hardware y software.

De este modo, cada vista es una abstracción que se centra en un aspecto de la especificación del sistema completo. Las vistas no deben considerarse capas o métodos de diseño, sino proyecciones de cierto conjunto de aspectos. Por otro lado, las vistas no son independientes ya que representan una parte del sistema completo, y existirán restricciones comunes en la especificación de distintas vistas debido a la relación entre las entidades reales representadas en cada una y a que la misma entidad puede estar incluida en varias vistas. Así, por ejemplo, puede que sea necesario hacer algunas consideraciones en las vistas de empresa o de información debidas a elecciones de distribución llevadas a cabo en las vistas computacional, de ingeniería y de tecnología.

El modelo de referencia de ODP es genérico, pero para dominios más específicos puede ser necesario refinarlo y especializarlo para acomodarlo a las necesidades particulares dando como resultado:

- Modelos de referencia específicos, que cubren dominios determinados y que utilizan los conceptos y funciones comunes dados en el modelo de referencia general y otros adicionales específicos del dominio. Así por ejemplo se ha desarrollado TINA (Telecommunication Information Networking Architecture) [133] y se está desarrollando HISA (Healthcare information systems architecture) [8,44,45].
- La estandarización de funciones específicas necesitadas en dominios determinados e identificadas en un modelo de referencia específico. Por ejemplo las interfaces necesarias en una conexión telefónica o en la identificación de una persona.

4.3.4. Algunos conceptos importantes en ODP

En este punto se presentan algunos de los conceptos manejados en ODP e imprescindibles para entender la filosofía de esta recomendación.

Objetos

Las especificaciones de sistemas ODP se expresan en términos de objetos. Un objeto contiene información y ofrece servicios. Un sistema se compone de objetos interactuantes. Un objeto se caracteriza por la encapsulación, la abstracción y el comportamiento.

La encapsulación es la propiedad por la cual la información contenida en un objeto es accesible solamente mediante interacciones en las interfaces soportadas por el objeto. La abstracción implica que los detalles internos de un objeto están ocultos con respecto a otros objetos y es de capital importancia para el tratamiento de la heterogeneidad, pues permite implementar servicios diferentes de diversos modos, utilizando mecanismos y tecnologías distintas, lo que hace posible la portabilidad y la interoperatividad.

La abstracción proporciona también una buena separación entre los objetos, lo que permite reemplazarlos o modificarlos sin cambiar su entorno, a condición de que continúen soportando los servicios esperados. Un modelo de objeto proporciona modularidad y la aptitud para componer nuevos módulos a partir de módulos existentes: estas capacidades son importantes para la construcción de sistemas flexibles y promueven la reutilización para acrecentar la productividad.

El término objeto puede ser considerado equivalente a otros términos ampliamente utilizados en computación distribuida como son componente o agente; todos ellos pueden considerarse como elementos computacionales que se podrán relacionar con otros a través de interfaces. Estos elementos serán reactivos o proactivos, es decir que responderán a llamadas y eventos externos, o generarán llamadas a otros componentes cuando, de forma autónoma, lo consideren necesario.

Interfaces

Los objetos sólo pueden interactuar a través de interfaces. Una interfaz representa una parte del comportamiento del objeto de manera que cada interfaz está identificada por un conjunto de interacciones en las que el agente puede participar. Una interfaz existe en un punto de interacción y un punto de interacción puede presentar varias interfaces.

Las interacciones entre agentes no están restringidas y pueden incluir, por ejemplo, interacciones asíncronas e interacciones síncronas multidireccionales. Pueden clasificarse en:

- **operaciones**, que son similares a procedimientos y son invocadas en interfaces designadas.
- **flujos**, que son abstracciones de secuencias continuas de datos entre interfaces.
- **señales**, que son interacciones atómicas elementales.

Las operaciones reflejan el paradigma cliente/servidor. Una operación es una interacción en la que un objeto, que juega el papel de cliente, invoca a un objeto, que asume el rol de servidor, para que realice alguna acción. Hay dos tipos de operaciones:

- Una **interrogación**, en la que el servidor entrega una respuesta a la petición del cliente.
- Un **anuncio**, en el que no hay respuesta a la petición del cliente.

Un flujo se caracteriza por su nombre y su tipo, que especifica la naturaleza y el formato de los datos intercambiados. En una relación de este tipo un objeto es el productor del flujo y el otro el consumidor. La semántica exacta de los flujos se deja sin definir en un modelo computacional. De hecho, puede haber muchas semánticas diferentes para flujos, que dependerán del dominio de aplicación.

Las señales constituyen el nivel más bajo de descripción de interacciones entre objetos. Una señal es una acción atómica compartida, por pares, y que produce una comunicación unidireccional de un objeto computacional iniciador a un objeto respondedor.

La figura 6 presenta un resumen de los tres tipos de interacciones y la representación de las mismas utilizada en esta Tesis Doctoral.

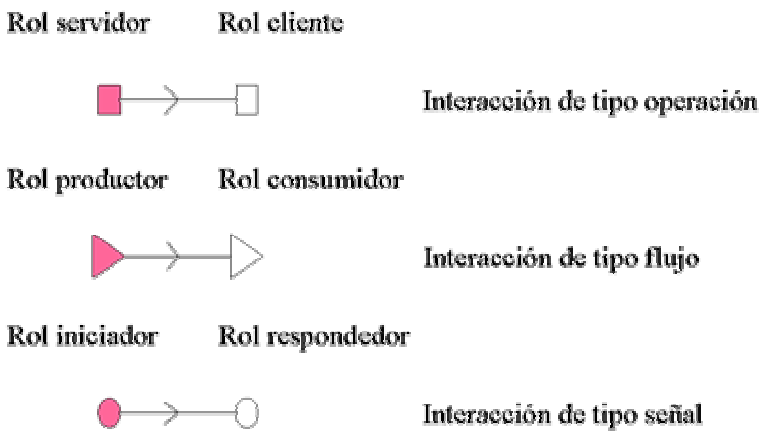


Figura 6. Tipos de interacciones en ODP

Una interfaz de un objeto o agente se caracteriza por una firma, un comportamiento, y un contrato de entorno. La firma define el conjunto de operaciones (interfaz de operación), flujos (interfaz de tren) o señales (interfaz de señal) soportadas en la interfaz y el papel que el objeto juega en cada interacción. De este modo, por ejemplo, en una misma interfaz se pueden definir varias operaciones de manera que en unas el objeto asuma el papel de cliente y en otras el de servidor.

El comportamiento se describe por las secuencias autorizadas de interacciones del objeto que están asociadas con la interfaz. Puede incluir acciones internas del agente y estará restringido por el entorno del mismo, en particular por las interacciones con otras interfaces.

Por último, cada especificación de interfaz contiene también un contrato de entorno que indica un conjunto de restricciones de calidad de servicio impuestas a un agente y su entorno (otros objetos e infraestructura de soporte). Estos requisitos pueden incluir factores como, por ejemplo, control de tiempos, disponibilidad, fiabilidad o la provisión de cierta tolerancia a las averías. Si el entorno proporciona el nivel requerido de calidad de servicio está entonces garantizado, por diseño, que el objeto proporciona un cierto nivel de calidad de servicio. Por tanto la calidad de servicio especificada para una interfaz expresa por un lado los requisitos impuestos a su entorno, y por otro la calidad de servicio mostrada por el objeto en un entorno que satisface los requisitos.

En consecuencia, será necesario especificar las firmas, el contrato de entorno y el comportamiento de las interfaces de un objeto para que éste se considere completamente especificado.

Vinculaciones

Las interacciones entre interfaces sólo son posibles si se ha establecido entre ellas alguna vinculación, es decir, algún trayecto de comunicación. Las acciones de vinculación pueden ser primitivas o compuestas.

Una acción de vinculación primitiva, como la mostrada en la figura 7, permite la vinculación de dos interfaces de dos objetos computacionales iguales o diferentes. Se requiere que las dos interfaces que intervienen sean del mismo tipo (operación, tren o señal), y que tengan un tipo de firma y roles complementarios. Una acción de vinculación primitiva es ejecutada por uno de los dos objetos que intervienen y tiene por efecto configurar cada interfaz para que se produzcan las interacciones definidas.

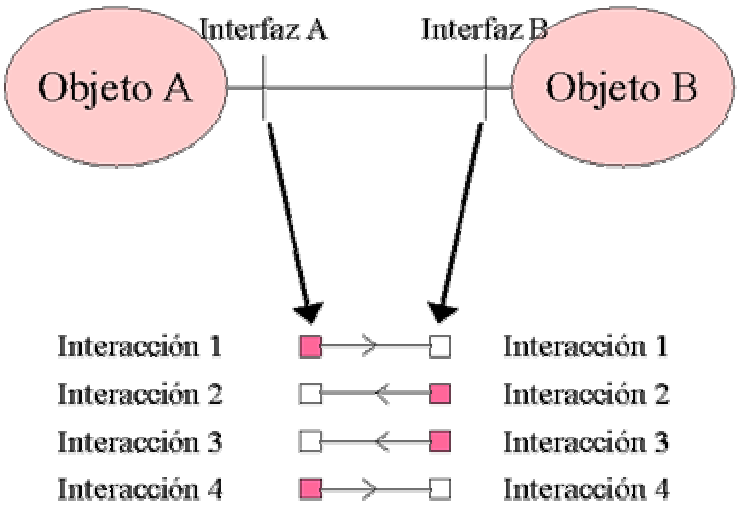


Figura 7. Vinculación primitiva

Una acción de vinculación compuesta, como la mostrada en la figura 8, permite la vinculación de dos o más interfaces del mismo tipo o de tipos diferentes por medio de un objeto de vinculación. Este objeto es instanciado por el que ejecuta la acción de vinculación, que puede intervenir o no en la misma, y presenta el conjunto apropiado de interfaces sobre las que ejecuta acciones de vinculación primitiva para conectarlas a los objetos que participan en la vinculación compuesta. Instancia también un conjunto de interfaces de control a través de las cuales podrán ser controladas sus operaciones. El comportamiento del objeto de vinculación es el que refleja la semántica de comunicación y al no estar limitados por el modelo computacional los tipos de objetos de vinculación, las estructuras de comunicación entre objetos son ilimitadas.

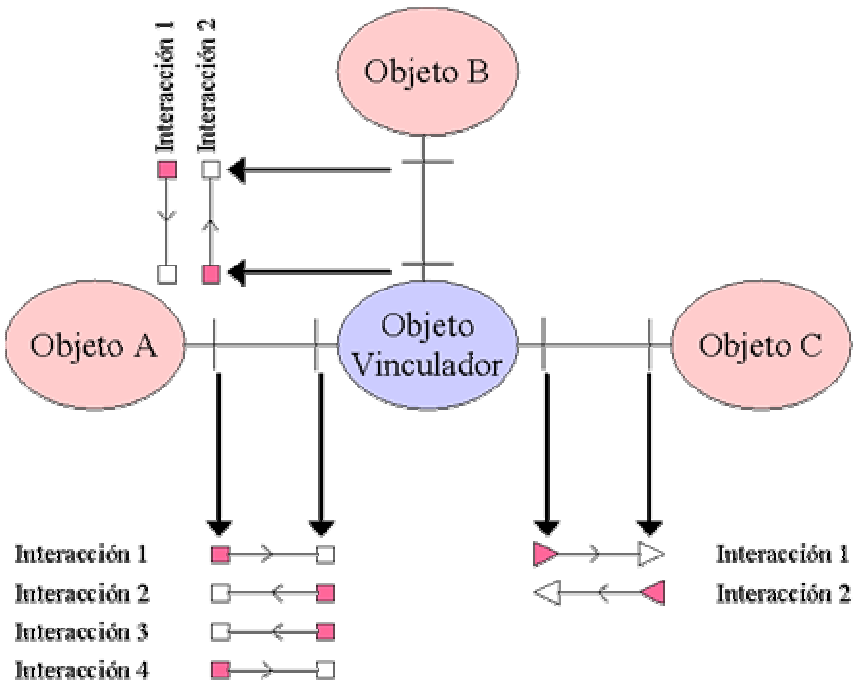


Figura 8. Vinculación compuesta

Por tanto, un agente software u objeto implementa una o varias interfaces. Dentro de cada una de las interfaces se define un conjunto de interacciones del mismo tipo, un comportamiento y un contrato de entorno. Para cada una de las interacciones el agente asume un determinado papel. A través de cada una de estas interfaces el agente puede interaccionar con otro agente que implementará la interfaz complementaria, mismas interacciones con roles complementarios, cuando se establece un vínculo entre ellos. Un ejemplo de vínculo entre interfaces, en este caso de tipo operación, se muestra en la figura 7.

Otro aspecto importante en la arquitectura es que existe la posibilidad de realizar composiciones de agentes. La figura 9 representa esta idea. Esto permite organizar la especificación de un sistema distribuido como un conjunto de especificaciones, cada una de las cuales trata un nivel diferente de abstracción, y en el proceso de diseño de una aplicación distribuida se puede utilizar la descomposición y la composición para organizar el ciclo de diseño en sucesivos pasos, de distintos niveles de abstracción.

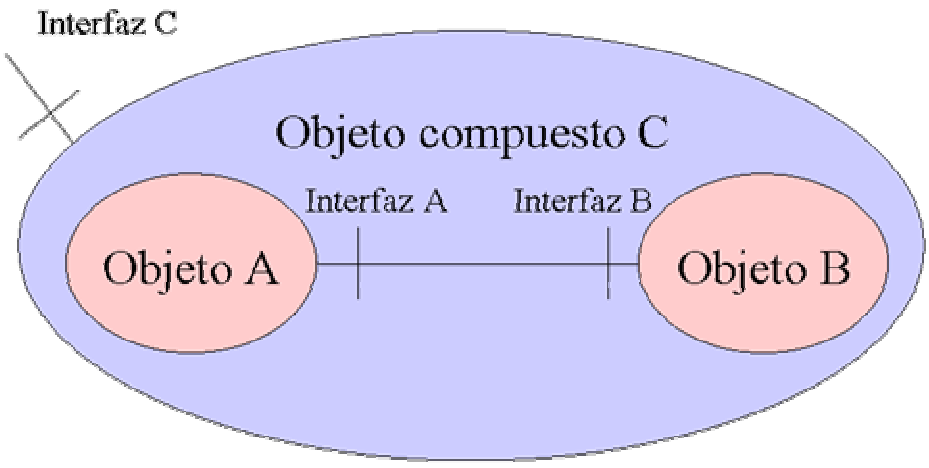


Figura 9. Objeto compuesto a partir de la interacción de otros

Por tanto la composición es una herramienta de modelado muy potente que permite que un subsistema sea tratado como un solo objeto de nivel superior ocultando las interfaces entre los objetos de nivel inferior que lo componen. Facilita que una aplicación distribuida compleja se descomponga en un número de objetos más simples que a su vez podrían descomponerse a un nivel menor de abstracción. Así, la composición proporciona un mecanismo de descripción de las relaciones jerárquicas entre un objeto compuesto y sus partes, facilitando una descripción jerárquica de una aplicación distribuida

Servicio

El concepto Servicio se entiende como [126] **"el conjunto de capacidades proporcionadas por un objeto en una interfaz computacional. Un servicio es una instancia de un tipo de servicio ofrecido en una interfaz."**. Esta definición no implica que el paradigma de comunicación entre el agente que ofrece capacidades y el que las utiliza sea, necesariamente, del tipo cliente/servidor. Las prestaciones ofrecidas por un componente pueden ser llevadas a cabo por composición, gracias a la interacción del agente con otros a través de interfaces en las que el tipo de interacción y el papel del objeto pueden ser cualquiera. Así se podría distinguir entre servicios básicos y servicios compuestos, como se muestra en la figura 10.

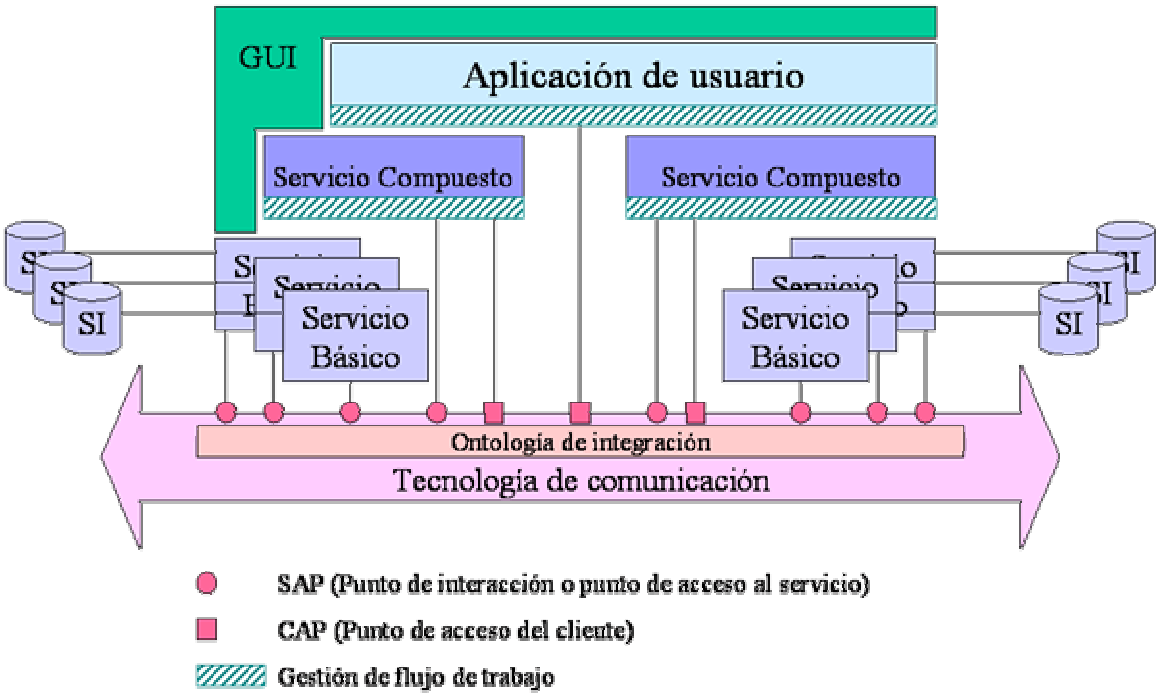


Figura 10. Relación entre servicios básicos, compuestos y aplicaciones

Federación

En ODP se entiende federación como un tipo particular de comunidad en la que sus miembros provienen de grupos pertenecientes a distintas autoridades o dominios y que cooperan para llevar a cabo algún objetivo.

La evolución de los sistemas distribuidos tiende cada vez más hacia la integración de subsistemas gestionados autónomamente que quieren compartir información para llevar a cabo transacciones de interés común. De manera que la creación de federaciones y la expresión de las reglas que las gobiernan, incluidas las políticas de interfuncionamiento entre dominios, son fundamentales en la especificación del sistema y proporcionan la modularidad que permite la escalabilidad del sistema. Estos mecanismos se recogen en la vista de empresa y estarán relacionados con la especificación de interceptores entre dominios, que pueden ser administrativos (por ejemplo si utilizan distintos procedimientos de seguridad o gestión), tecnológicos, o ambos, y se desarrollan en la vista de ingeniería.

4.4. Una especialización de ODP: la norma HISA

4.4.1. Introducción

Actualmente el CEN está desarrollando una nueva prenorma, la prEN 12967 [8,44,45] (Health informatics – Service architecture), donde se describe una metodología para el diseño de arquitecturas de sistemas distribuidos en entornos sanitarios, Healthcare Information System Architecture (HISA). Esta norma sustituye a la anterior ENV 12967-1 (Medical informatics –- Part 1: Healthcare middleware layer), de 1997, incluyendo algunas modificaciones. HISA impone que para la normalización de sistemas abiertos y distribuidos en el entorno sanitario éstos deben estar modelados según el estándar ODP (Open Distributed Processing). De manera que está estructurada en tres partes que se corresponden con las 3 primeras vistas definidas en ODP:

- Parte 1: *Enterprise viewpoint*
- Parte 2: *Information viewpoint*
- Parte 3: *Computational viewpoint*

En el mes de octubre de 2005 se han publicado, en la zona de acceso libre del portal del CEN [37], nuevas versiones de las partes 1 y 2. Respecto a la parte 3 sólo hay publicados borradores muy inmaduros.

Esta prenorma se comenzó a desarrollar al analizar que actualmente una de las mayores necesidades en las organizaciones sanitarias es la integración y la interoperatividad de las aplicaciones ya existentes, como ya se viene señalando en este trabajo. Los mecanismos de integración deben asegurar que el suministro de servicios no se interrumpa durante estos procesos y facilitar la migración de sistemas propietarios y monolíticos hacia la idea de sistemas abiertos y modulares. El coste de estas soluciones de integración representa también un aspecto crucial que es necesario evaluar con detenimiento. Estos objetivos pueden ser llevados a cabo con la planificación de arquitecturas abiertas, basadas en plataformas de integración independientes de las aplicaciones a integrar. Todos los aspectos de la estructura sanitaria deben ser cubiertos por la arquitectura, que debe considerar la información de interés común a las distintas aplicaciones y la lógica de negocio, estructurándolos de acuerdo a criterios y paradigmas independientes de aspectos sectoriales específicos, requisitos temporales o soluciones tecnológicas.

Basándose en estas consideraciones se especifica esta norma, cuyo objetivo es dar una guía para la descripción, planificación y desarrollo de nuevos sistemas, así como para la integración de sistemas ya existentes, basada en una arquitectura que integre los datos comunes y la lógica de negocio de forma independiente a las aplicaciones individuales y permitiendo la accesibilidad a toda la información del sistema a través de servicios. Esta arquitectura estará soportada por alguna plataforma que proveerá de las necesidades de comunicación y coordinación de componentes, como se muestra en la figura 11, extendiendo la misma para soportar los requisitos particulares del dominio sanitario. De manera que cuanto más avanzados sean los servicios ofrecidos por la tecnología subyacente menores serán los esfuerzos a realizar en la arquitectura del dominio sanitario.

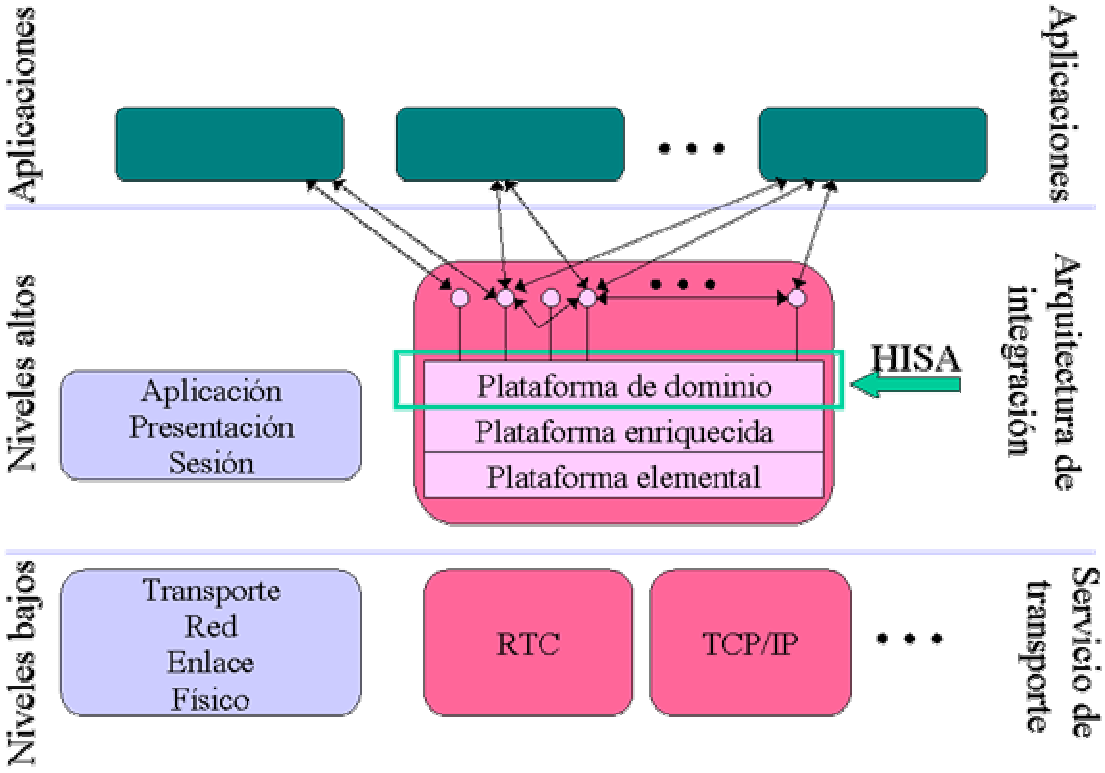


Figura 11. Papel de HISA en la arquitectura sanitaria

La figura 11 muestra el lugar que HISA ocuparía dentro de una arquitectura de sistemas sanitarios. Esta norma debe aportar soluciones a los requisitos propios del dominio sanitario que no estén cubiertos por la arquitectura de integración, o *middleware*, elegida por la organización. Por otro lado las particularidades del servicio de transporte

serán totalmente transparentes a HISA ya que estas funciones se resuelven en las capas más bajas del modelo.

La prenorma se planifica con dos propósitos fundamentales:

- Identificar una metodología que describa los sistemas de información sanitarios utilizando lenguajes, notaciones y paradigmas que faciliten la planificación, diseño y comparación de sistemas.
- Identificar los aspectos fundamentales de arquitecturas abiertas que permitan la integración e interoperatividad de sistemas de información sanitarios.

Esta norma no sustituye a otros estándares, como GPICS [34,42,43], prEN13606 [27,28,39-41] o DICOM [49], sino que los complementa con la finalidad de conseguir la integración de distintos sistemas de información sanitarios.

El estándar no pretende normalizar el conjunto completo de especificaciones necesarias en un sistema sanitario distribuido. Por el contrario sólo formaliza los aspectos fundamentales que han sido identificados como comunes en todos los países europeos y que se consideran esenciales en cualquier sistema de información sanitario actual. Estas especificaciones se formalizan evitando cualquier dependencia de tecnologías, productos o soluciones específicas. El hecho de no incluir las vistas de ingeniería y tecnológica así lo manifiesta. Los sistemas conformes a HISA deberán ser descritos en términos de los tres puntos de vista normalizados y complementados con las especificaciones de los aspectos técnicos y de infraestructura necesarios. Estas características se describirán cumpliendo los criterios definidos por ODP para las vistas de ingeniería y tecnología.

4.4.2. Vista de empresa

La primera parte de este estándar, *HISA Enterprise viewpoint* [8], proporciona una especificación detallada de los requisitos de empresa que debe satisfacer la arquitectura de integración. Utilizando un alto nivel de abstracción, es decir sin considerar muchos de los detalles de los sistemas sanitarios, en el paradigma de “Empresa sanitaria” se pueden identificar tres flujos de trabajo fundamentales durante las actividades de los usuarios del sistema:

Flujo de sujeto de cuidado (centrado en el paciente)

Se refiere a las actividades de los usuarios para gestionar la información personal y estadística relativa a un paciente (o sujeto de cuidado, que es un concepto más amplio) y a la gestión de encuentros entre éste y las organizaciones sanitarias implicadas en la arquitectura. También se consideran parte de este flujo de trabajo todas las actividades relativas a la gestión de facturación por los servicios prestados al sujeto de cuidado.

Flujo de gestión de actividad (centrado en el cuidado)

Se refiere a las actividades de los usuarios para gestionar los distintos tipos de actividades que pueden ser ejecutadas en la organización durante su ciclo de vida completo. Se incluyen los aspectos relativos, por ejemplo, a la petición inicial, el registro, la planificación, la ejecución y la emisión de informes.

Flujo de información clínica (centrado en la información)

Este flujo se refiere a las actividades de los usuarios para gestionar los datos clínicos, incluidos, por ejemplo, aspectos relativos a la recuperación y validación, así como a la agregación y estructuración de datos de acuerdo a los requisitos específicos de distintos usuarios y disciplinas.

Otras actividades

También deberán existir servicios en la arquitectura que den soporten a las siguientes actividades de usuario:

- Gestión de información relativa a la descripción de la estructura de la organización, los usuarios del sistema y los criterios de autorización que deben cumplir éstos para que se les permita el acceso a los datos y la ejecución de las prestaciones ofrecidas.
- Gestión de los recursos disponibles en la organización y de las reglas y criterios que deben cumplir para que puedan estar disponibles y ser utilizados.
- Gestión de clasificaciones, criterios de codificación y diccionarios utilizados en distintos sectores para la clasificación de la información gestionada.

La arquitectura debe proveer de servicios que permitan la estructuración de datos y la interacción con otros sistemas de información a través de mecanismos basados en mensajes y otros formalismos conformes a estándares de comunicación.

4.4.3. Vista de información

La segunda parte del estándar, *HISA Information viewpoint* [44], proporciona la especificación detallada del modelo de información que debe ser implementado en la plataforma de integración. De acuerdo a los requisitos identificados en la vista de empresa, la vista de información debe contener siete grupos de objetos, cada uno de ellos organiza y almacena la información necesaria en las actividades identificadas en la vista de empresa. Por ello el nombre de cada uno de estos grupos de objetos refleja las actividades a las que corresponden

- Objetos del sujeto de cuidado
- Objetos de gestión de actividad
- Objetos de información clínica
- Objetos para la organización, usuarios y autorizaciones.
- Objetos de recursos
- Objetos de clasificación
- Objetos de mensajes

4.4.4. Vista computacional

La última parte del estándar, *HISA Computational viewpoint* [45], proporciona una especificación detallada de los servicios que deben proporcionarse en la arquitectura.

Como se mostraba en la figura 10 se identifican dos tipos de servicios: básicos y compuestos.

Servicios básicos

Permiten recuperar y manipular instancias de objetos pertenecientes al modelo de información. Representan los servicios fundamentales de la arquitectura, imprescindibles para:

- Ser utilizados directamente por las aplicaciones y permitirles manipular la información para sus propósitos específicos.
- Ser utilizados como bloques básicos de servicios compuestos.

- Asegurar el principio fundamental de sistema abierto y la propiedad del cliente sobre la información.

Servicios compuestos

Implementan las transacciones de negocio completas, relativas a las actividades de usuario disponibles, implicando a múltiples objetos y datos según las reglas específicas de la organización.

Este estándar, hasta la fecha, sólo identifica un conjunto mínimo de requisitos, fundamentales y comunes a todas las organizaciones sanitarias y que deben ser satisfechos por la plataforma de integración en términos de las actividades de empresa soportadas, la información gestionada y los servicios provistos. Pero la especificación del estándar debe ser extensible a lo largo del tiempo de acuerdo con la evolución de las iniciativas de estandarización. Así, los productos que se desarrollen conformes a este estándar deben permitir extensiones para satisfacer los requisitos nacionales y locales que vayan surgiendo. La metodología para formalizar estas extensiones está también definida por la norma y estas extensiones deben garantizar la compatibilidad y consistencia con el estándar, en el sentido de que los principios ya definidos y los servicios ya formalizados no deben ser cambiados, ni en cuanto al objetivo ni en cuanto a la interfaz.

4.4.5. Conformidad

En cuanto a los criterios de conformidad, se provee de una metodología para describir los sistemas de información sanitarios y la especificación de una plataforma de integración capaz de soportar el sistema completo de la organización sanitaria. Se distinguen dos tipos de conformidad.

Conformidad de los documentos de la especificación con la metodología HISA.

Un sistema de información definido conforme a HISA debe ser conforme a la metodología definida y debe consistir en los siguientes documentos: El paradigma estratégico y las vistas de empresa, de información y computacional.

El paradigma estratégico es un documento muy conciso que identifica, con un alto grado de abstracción, los requisitos y objetivos estratégicos del sistema. Estará descrito en lenguaje natural y debe contener:

- El objetivo del sistema de información dentro del conjunto de la empresa.
- Los procesos de organización fundamentales, en forma de términos y definiciones, que pueden identificarse en la empresa y que pueden ser importantes para el sistema.
- Las restricciones y objetivos fundamentales que deben ser satisfechos.
- Evolucionando y refinando este paradigma estratégico, y siempre conforme a él, la arquitectura debe ser descrita a través de las distintas vistas: de empresa, de información y computacional.

Se recomienda además que para complementar estos documentos se describa también la implementación física desde los puntos de vista de ingeniería y tecnología. La empresa que cree estos documentos puede hacerlo individualmente, pero puede también hacerlo en cooperación con un suministrador de productos HISA o consultoras especializadas en HISA.

Conformidad de productos middleware a los requisitos de la arquitectura HISA

Los productos conformes a la norma deben:

- Estar descritos según la metodología y documentos definidos en la norma.
- Implementar la arquitectura conforme a los requisitos definidos en la norma.
- Implementar un modelo de información con todos los objetos y conforme a los requisitos definidos en la vista de información.
- Proporcionar al menos todos los servicios definidos en la vista computacional.
- Ser extensible y acomodar requisitos locales y nuevos estándares de acuerdo a los criterios de extensión de la norma.

El sistema debe formalizar los mecanismos para soportar los estándares CEN relativos a la historia clínica y sus implementaciones nacionales, así como los principales estándares de mensajería y comunicación definidos en los escenarios internacionales de acuerdo a los requisitos de la organización.

4.4.6. Conclusiones

Como se ha venido presentando, las características de heterogeneidad de los sistemas de información sanitarios, desarrollados independientemente por distintos proveedores que pueden basar sus soluciones en distintos estándares o realizarlas de forma completamente propietaria, hacen que la integración de sistemas de información en este entorno pueda entenderse como un problema particular de computación distribuida. De manera que la elección de ODP como método de modelado para la especificación de estándares se entiende como una elección muy acertada por parte del CEN ya que puede facilitar una arquitectura independiente de la tecnología subyacente.

Dado que el paradigma de integración más idóneo en este entorno es el federado, la norma debería hacer hincapié en las necesidades de estos mecanismos de interrelación. El planteamiento actual de HISA no facilita suficientemente las tareas en ese sentido, ya que obliga a adoptar modelos de información y computacionales muy estrictos.

Por otro lado, si el modelo de distribución propuesto pretende ser una base firme para la evaluación de la conformidad de componentes respecto al mismo, debe definir muy claramente los requisitos que va a solicitar de los implementadores y los mecanismos que se van utilizar para realizar las pruebas de conformidad. Las pruebas de conformidad de los sistemas distribuidos son potencialmente más complejas y costosas que en los protocolos de comunicación tradicionales. Muchas de las especificaciones son muy abstractas, ya que se basan en patrones de comportamiento orientados a la aplicación, sin considerar el mecanismo de comunicación que se usará. Validar la conformidad en estas circunstancias requiere una interpretación muy estudiada de los resultados de las pruebas antes de alcanzar un veredicto. Incluso puede ser necesaria alguna técnica de validación distinta a las tradicionales.

El modelo de referencia incluye una metodología básica para evaluar la conformidad con determinada arquitectura, y la forma en la que las pruebas de conformidad deben ser llevadas a cabo en cada vista. Este riguroso tratamiento de la conformidad es la base de los sistemas abiertos. Como vimos anteriormente, uno de los grandes problemas del desarrollo de estándares es precisamente la dificultad en la evaluación de conformidad, tanto que actualmente están teniendo un gran auge iniciativas como IHE que surgen principalmente de la necesidad de mejorar este aspecto.

La especificación de los servicios debe depender lo menos posible de la tecnología, y por eso HISA sólo especifica las tres primeras vistas. Sin embargo, la interrelación existente entre vistas puede obligar a hacer consideraciones relativas a la tecnología seleccionada. Algunos de los requisitos que se imponen en HISA podrían estar cubiertos por las plataformas de integración subyacentes.

Capítulo 5. Servicio demográfico

5.1. Introducción

En este capítulo se analizan las tareas e importancia de los servidores demográficos en un sistema sanitario. Se presentan algunos de los principales modelos de datos demográficos normalizados y se describe el principal modelo de servicio de identificación de personas utilizado en el contexto sanitario: el servicio de identificación de personas [29] del grupo CORBAmed del OMG [52]. Para terminar se realiza una discusión de los aspectos estudiados analizando puntos fuertes y débiles, en los que esta Tesis Doctoral realiza diferentes aportaciones.

5.1.1. Los servicios demográficos

Un servidor demográfico debe realizar dos tareas fundamentales: la identificación inequívoca de personas y la gestión de información demográfica sobre las mismas. Será ampliamente utilizado por componentes que se encarguen de la recuperación de datos clínicos o de otras tareas tales como las relativas a la seguridad del sistema. Los agentes demográficos podrían extenderse para permitir, basándose en las mismas interfaces o con pocas modificaciones, la gestión de otras entidades, como por ejemplo organizaciones.

Los datos almacenados en cualquier sistema de información clínico pueden recuperarse conforme a distintas vistas. La Historia Clínica (HC), por ejemplo, es una vista centrada en el paciente, en la que se recupera toda la información almacenada relativa a determinada persona. Por tanto una de las principales características de la HC es que el paciente es el eje central, de manera que la identificación y recuperación de datos personales sobre el mismo es crucial para afrontar el desarrollo de la HC Electrónica (HCE). Sin embargo la información demográfica de pacientes no es la única manejada en la HC, también puede incluirse información sobre la organización o el personal que atiende al paciente, familiares del mismo, etc.

Por otro lado, la información demográfica no sólo se utiliza en la HC, sino en muchas otras actividades de una organización sanitaria, como la gestión de nóminas y recursos

de la organización, o incluso la extracción de información clínica para realizar, por ejemplo, estudios epidemiológicos.

5.1.2. Identificación de personas

Hay numerosas actividades en la organización sanitaria que requieren la identificación unívoca de personas. Por ejemplo, el primer paso para recuperar la HCE relativa a un paciente es, precisamente, identificar al mismo de forma inequívoca [23] y, por motivos de seguridad, también a la persona que inicia esta sesión de consulta. Esta identificación devolverá como resultado un identificador único dentro del sistema que, con total seguridad, no se va a repetir para dos personas distintas. Una correcta gestión de estos identificadores de personas (IP) es una tarea clave en la administración de un servidor demográfico.

Puede utilizarse como IP algún identificador gestionado por la administración a la que corresponde el sistema de información. En un sistema de información clínico español podría utilizarse, por ejemplo, el DNI, el pasaporte, el número de afiliación a la seguridad social, o el número de tarjeta sanitaria. Esta elección puede tener varios problemas:

- La necesidad de atender y registrar a alguien que no tenga asignado ese identificador. Por ejemplo, los menores de 14 años o los extranjeros no tienen DNI.
- Errores en la asignación de identificadores por parte de la administración. Por ejemplo, la asignación del mismo DNI a dos personas.
- Errores en la entrada de datos.

La solución a este conflicto es añadir al uso de estos identificadores un IP gestionado de forma local y autónoma en el sistema. Cuando se introduzca una persona nueva en el sistema éste le asigna un nuevo IP distinto a todos los demás, asegurándose que esta clave única no se duplica.

Pensar en utilizar únicamente este IP local olvidándonos de almacenar los identificadores administrativos es también un error. Es difícil conocer el identificador local asignado a una persona en el momento en el que se quiere acceder al sistema para recuperar la HCE, por lo que lo primero es buscar este IP. Para recuperar el identificador hay que partir de unos datos conocidos y fiables sobre la persona que se

pretende identificar; utilizar para esta búsqueda del IP local los identificadores administrativos, que sí son con alto grado de probabilidad conocidos, es una magnífica opción que asegurará que el número de candidatos que cumplen el criterio de búsqueda se reduzca notablemente. A partir de estos candidatos, si el usuario del servidor demográfico es directamente un agente humano la elección de la persona realmente buscada es relativamente sencilla; la mayor dificultad se encuentra en los procesos automatizados.

En el caso de varios sistemas trabajando de modo federado, la utilización de un IP local aumenta la complejidad de la identificación unívoca. Existe un IP local y único en cada uno de los sistemas, o dominios, pero éste habitualmente no coincide en los distintos sistemas de la federación, ya que se gestionan de forma completamente autónoma. El uso conjunto con IP administrativos facilitará esta tarea de identificación unívoca, ya que en cada dominio se podrá realizar una búsqueda más eficiente de los IP locales si ésta se hace a partir de IP administrativos.

5.1.3. Recuperación de información demográfica

Una vez que se conoce el identificador único de la persona en el sistema también es tarea del servidor demográfico proporcionar mecanismos para recuperar sus datos demográficos, aquellos que no son de índole sanitaria. Toda HCE y todo documento incluido en la misma, contiene los datos personales del paciente y, en la mayoría de las ocasiones, del personal facultativo responsable de cada tarea.

5.2. Modelos de datos

En cualquier servidor demográfico se deben utilizar modelos de información para los datos demográficos propiamente dichos, para los identificadores y para otras estructuras que sirvan de apoyo al desarrollo del servidor. A continuación se presentan algunos de los modelos de información demográfica y de identificación más importantes.

5.2.1. OpenEHR

El **modelo de referencia para datos demográficos** dado por el OpenEHR [33] está pensado para su utilización con arquetipos, siguiendo la filosofía del modelo dual ya presentada en el capítulo anterior de esta Tesis Doctoral. Este modelo de referencia, mostrado en la figura 12, establece los contenedores básicos para la información demográfica y deja en manos de cada implementación el desarrollo de arquetipos que

particularicen la información demográfica. La interoperatividad de sistemas se consigue al compartir el modelo de referencia y los arquetipos entre distintos dominios.

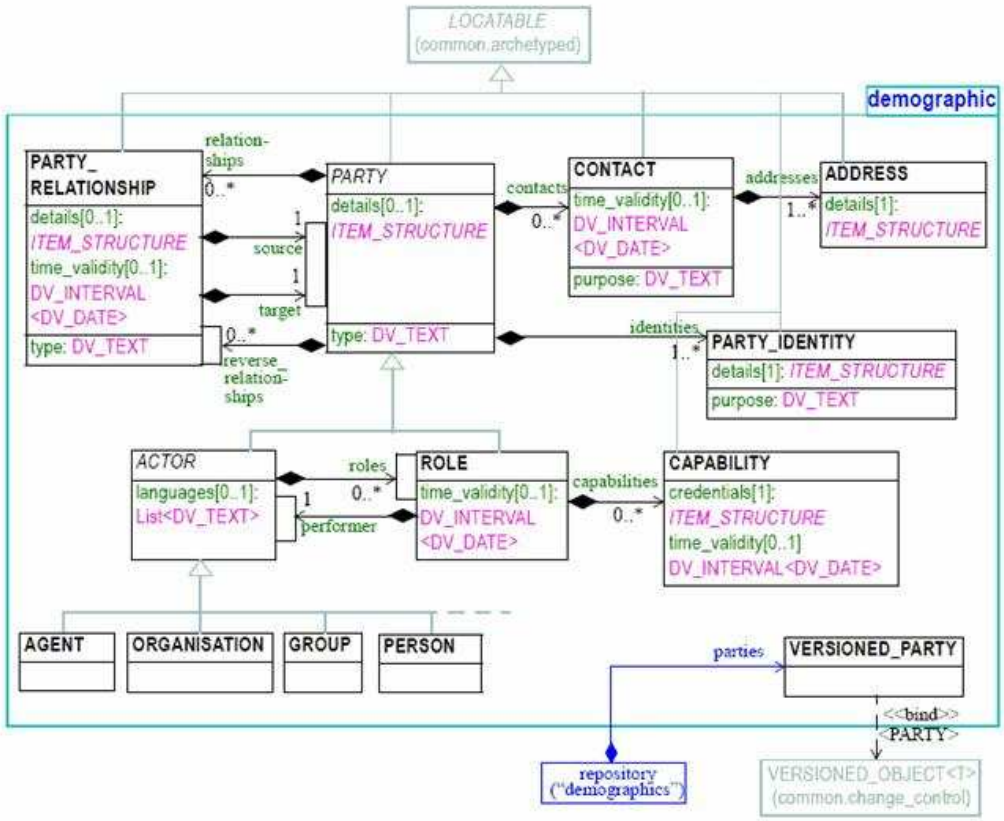


Figura 12. Modelo de información demográfico del OpenEHR (extraída de [33])

La clase que representa una persona, eje central de un servidor demográfico, se denomina **PERSON** y hereda de la clase **ACTOR**. Es de destacar que el modelo de referencia no se centra únicamente en personas sino que también permite representar grupos, organizaciones, agentes "no humanos" o cualquier otro actor implicado en la gestión de la HCE, y de los que sea necesario conocer datos que no son de naturaleza clínica. También establece un modo de representar el rol que determinado actor juega en la gestión de la HCE.

Identificadores

En cuanto a los identificadores en el modelo del openEHR se distingue entre el identificador del objeto que representa a una persona y el identificador de persona.

El identificador del objeto se encuentra en el atributo **uid** de **PARTY**, heredado de la clase **LOCATABLE**, que almacena un identificador unívoco para determinado objeto de

esa clase. Este atributo es del tipo **OBJECT_ID**, definido en el **modelo de referencia común** [66] dentro del paquete de identificadores, como muestra la figura 13. Esta clase modela un identificador de una entidad de información, no tiene que ser necesariamente una persona. Una entidad real puede corresponderse con varias entidades de información, por ejemplo puede almacenarse información de una persona (entidad real) en varios sistemas de información y cada una tendría un identificador distinto. Otra posibilidad es que hubiera varias versiones de una misma persona en el mismo sistema. La clase **OBJECT_VERSION_ID** puede utilizarse precisamente para distinguir identificadores que se refieren la misma entidad real pero los objetos que la representan tienen alguna característica que los distingue. Su uso sería, por ejemplo, para distinguir dos versiones, dentro de determinado sistema, de la información referente a la misma persona.

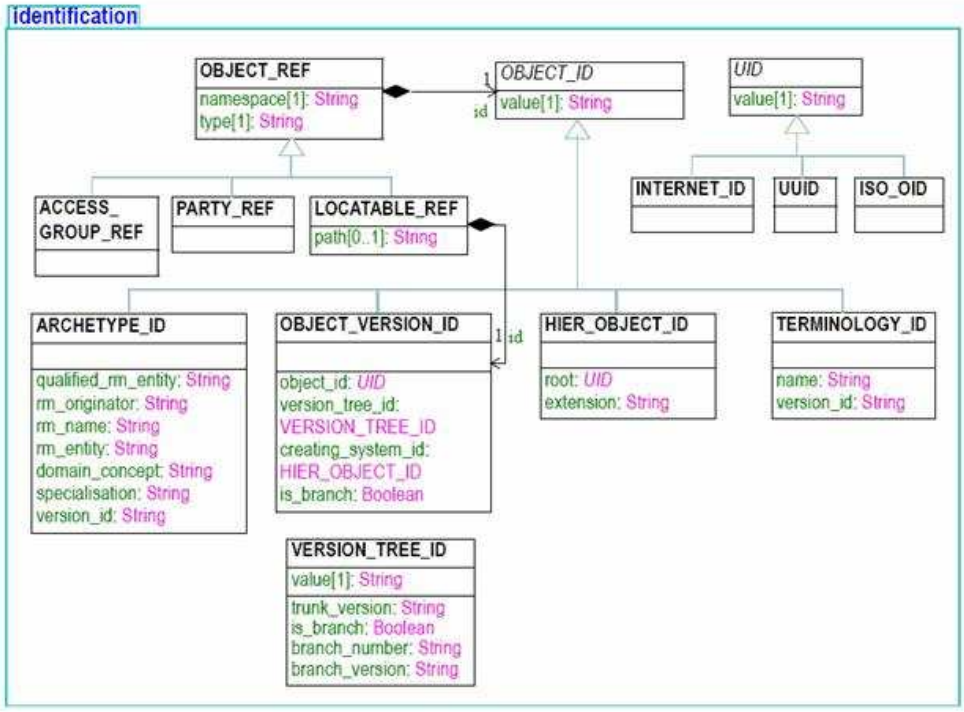
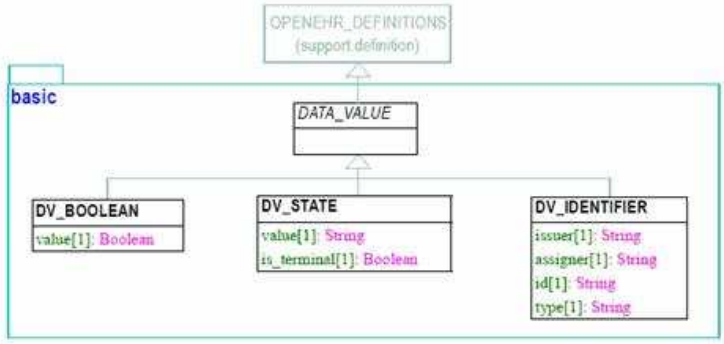


Figura 13. Modelo de información para identificadores en el openEHR (extraída de [66])

Para referirse externamente a un objeto que está disponible en determinado servicio, o también localmente, es necesario usar una instancia de **OBJECT_REF**, que proporciona una referencia tanto en entornos distribuidos como locales. Incluye el atributo

namespace que indica a quién pertenece o quién gestiona ese identificador, es decir que estará relacionado con el servidor que lo proporciona. De manera que en el caso que nos concierne el espacio de nombrado se referiría a un **servidor demográfico**. También existe un campo de tipo en el que se especifica el tipo concreto del objeto referenciado, que en el caso de referencias a personas sería el tipo **PERSON**. Y por supuesto se incluye el identificador que sería del tipo **OBJECT_ID** mostrado anteriormente.

Pero los identificadores de persona asignados por alguna administración son tratados en el modelo como un atributo más de la persona. La clase **PARTY** contiene una o varias identidades que representan las formas que pueden usarse para identificar a una persona, como puede verse en la figura 12. Aquí pueden incluirse distintos nombres, apodos y sería también un buen lugar para indicar los IP administrativos, como número de la seguridad social, DNI...y para el IP local. La clase **PARTY_IDENTITY**, que representa estas identidades, debe ser especificada a través de arquetipos, es decir que el modelo sólo especifica que debe tener un campo **details** del tipo **STRUCTURE** y que dependerá del sistema el formato exacto de esta estructura. Por otro lado openEHR también define un tipo específico para manejar identificadores y que podría ser usado cuando se creen los arquetipos para **PARTY_IDENTITY**, este tipo está representado en la figura 14.



CLASS	DV_IDENTIFIER	
Inherit	DATA_VALUE	
Attributes	Signature	Meaning
1..1	**issuer**: String	Authority which issues the kind of id used in the *id* field of this object.
1..1	**assigner**: String	Organisation that assigned the id to the item being identified.
1..1	**id**: String	The identifier value. Often structured, according to the definition of the issuing authority's rules.
1..1	**type**: String	The identifier type, such as "prescription", or "SSN". One day a controlled vocabulary might be possible for this.
Invariants	*issuer_valid*: issuer /= Void **and then not** issuer.is_empty *assigner_valid*: issuer /= Void **and then not** assigner.is_empty *id_valid*: id /= Void **and then not** id.is_empty *type_valid*: type /= Void **and then not** type.is_empty	

Figura 14. Tipo de dato para acomodar identificadores en el OpenEHR.

Datos demográficos

Se analizan ahora los contenedores que el modelo pone a nuestra disposición para almacenar información demográfica. La figura 12 muestra que el modelo es muy abierto, ya que está pensado para ser usado junto a arquetipos. La clase que representa a una persona, **PERSON**, hereda de la clase **ACTOR** y ésta, a su vez, de la clase abstracta **PARTY**. A continuación se analizan los atributos de la clase **PERSON**, donde se podrá acomodar la información demográfica, todos ellos heredados de estas dos clases.

PERSON puede incluir una o varias formas de contacto con esa persona. Un contacto está representado por la clase **CONTACT**. Esta clase incluye el intervalo de validez de esta forma de contacto y un conjunto de direcciones, de la clase **ADDRESS,** que representa direcciones geográficas o electrónicas. Las direcciones también serán especificadas en cada sistema según los arquetipos definidos.

Se incluyen en el modelo los idiomas que pueden usarse para comunicarse con esa persona y los papeles, o roles, que la persona puede tomar. Estos roles son de la clase **ROLE** que incluye el periodo de validez del rol y las credenciales o capacidades (clase

CAPABILITY) de la persona para actuar en ese papel. Estas credenciales también se especificarán según determinado arquetipo. Además se incluye un campo de periodo de validez para cada credencial.

5.2.2. CEN

Identificadores

Otra norma que sirve de referente para datos demográficos es el prenorma europea prEN 14822, Componentes de Propósito General para Mensajes (General Purpose Components for Messages, GPIC) [34,42,43]. Esta norma está divida en tres partes: introducción, componentes no clínicos y componentes clínicos. Dentro de la segunda parte, que establece los componentes para datos no clínicos, se incluye la información demográfica. Esta norma utiliza como tipos básicos los definidos en otro estándar europeo, Health informatics - Data types [134], algunos de los cuales se muestran en la figura 15. Los identificadores se especifican con el tipo **`Instance Identifier`**.

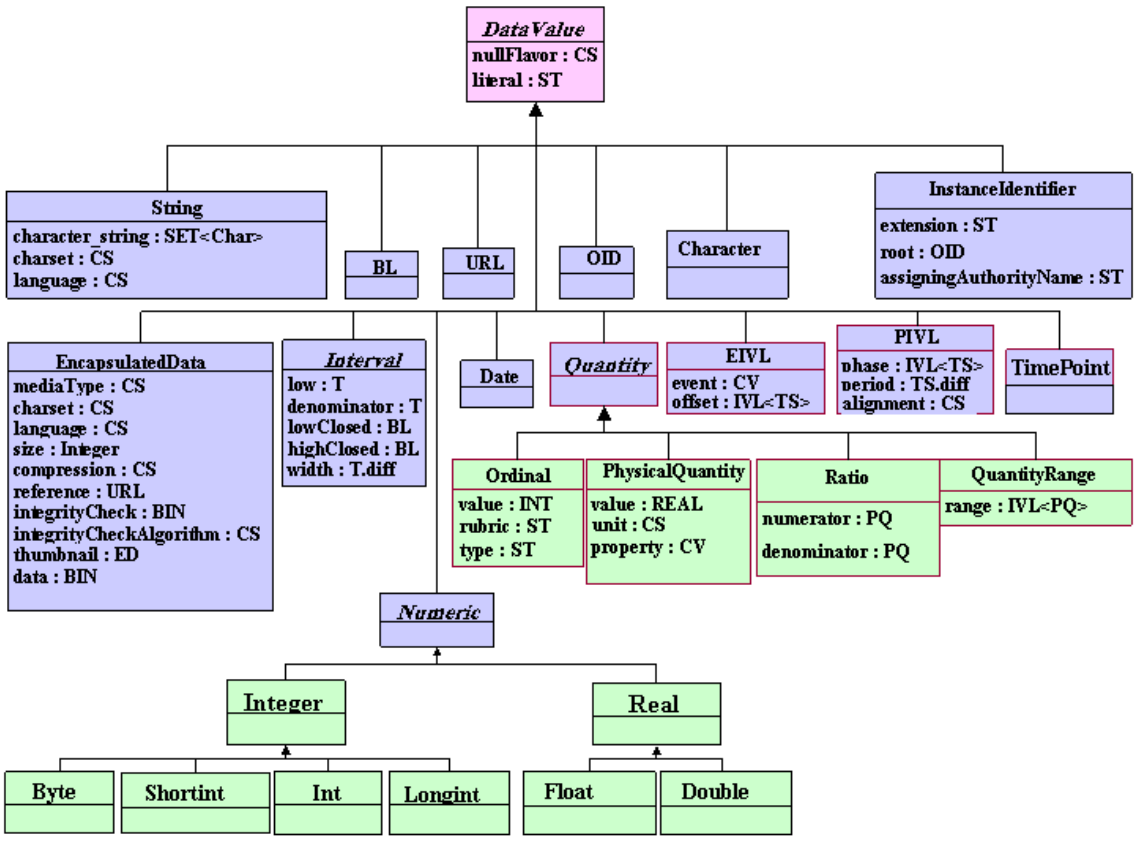


Figura 15. Tipos de Datos básicos dados por el CEN/TC251

Datos demográficos

El modelo de datos introducido proporciona una semántica mucho más rica en cuanto a este tipo de información se refiere. Es decir se detalla mucho más que las otras normas aquí expuestas la información demográfica de una persona. La ventaja de esta normalización más estricta es que facilita la interconexión de sistemas, a cambio de una mayor rigidez en la información intercambiada.

La figura 16 presenta algunos de los conceptos demográficos más interesantes de la norma.

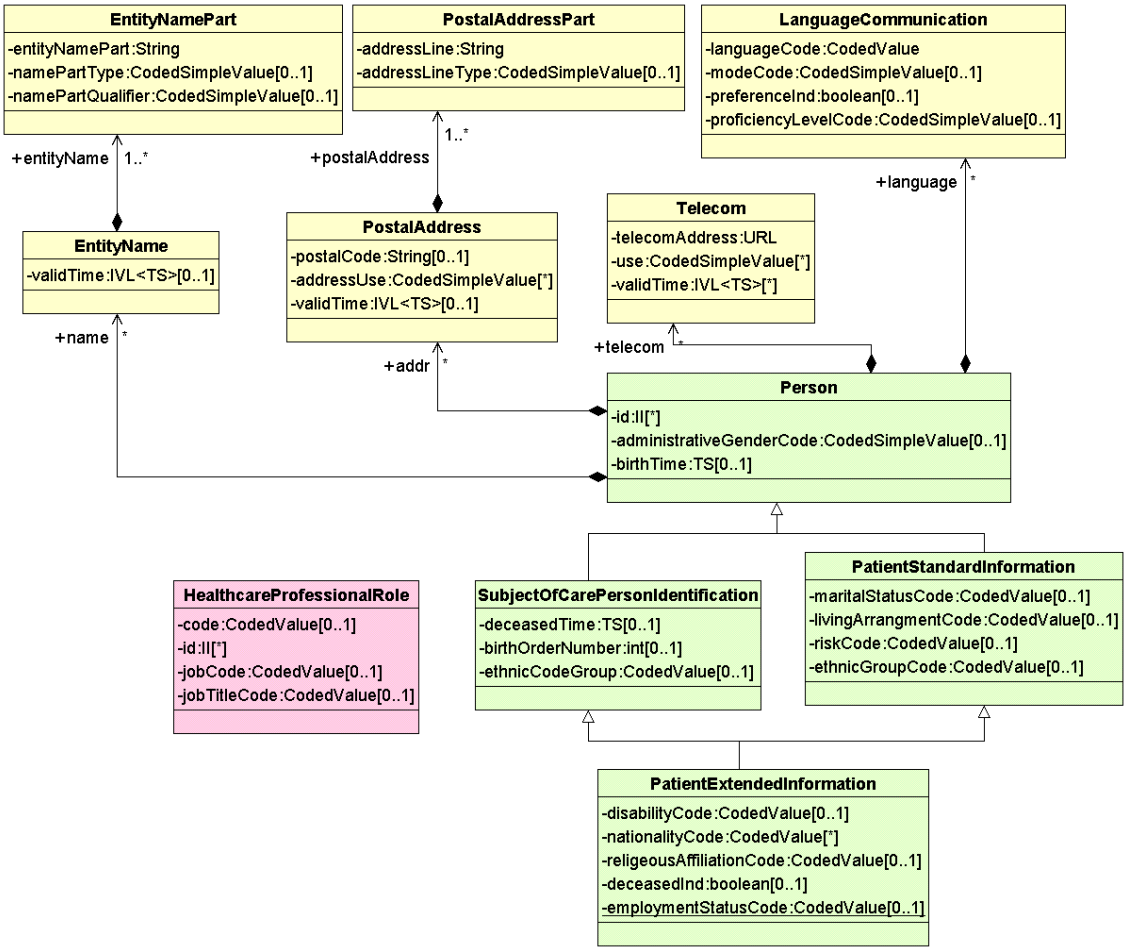


Figura 16. Resumen de información demográfica en los GPICS no clínicos

Como se puede observar, el nivel de detalle es mucho mayor que en el openEHR y por tanto la semántica está mucho más cerrada. Las clases seleccionadas para esta breve exposición han sido las relacionadas directamente con la clase **`Person`**, que representa una persona y contiene:

- Un conjunto de identificadores unívocos.

- Un conjunto de nombres por los que se conoce a la persona.
- El conjunto de direcciones postales.
- El conjunto de direcciones telemáticas.
- La fecha de nacimiento.
- Un código para indicar el sexo.
- El conjunto de idiomas con el que se puede comunicar con la persona.

En primer lugar un nombre de persona se especifica con una instancia de la clase **`EntityName`** que, como se puede comprobar, incluye un periodo de validez para ese nombre y una o varias partes de nombre o **`EntityNamePart`**. Una parte de nombre contiene, además del valor indicado con una cadena de caracteres, un campo que indica el tipo de elemento: si es un nombre propio, un apellido, un prefijo, etc, y otro que cualifica ese nombre indicando alguna propiedad, como si es el nombre de nacimiento, el profesional, etc.

En cuanto a las direcciones postales de una persona hay que utilizar objetos de la clase **`PostalAddress`**. Esta clase incluye el uso, o usos, que tiene esa dirección postal, por ejemplo si es la dirección de trabajo, de vacaciones, etc, y el periodo de validez de esa dirección. Además se incluye el campo de código postal y un conjunto de partes de dirección o **`PostalAddressPart.`** Cada parte de la dirección incluye el valor, expresado como una cadena de texto, y el tipo, que es un código que indica si se refiere a una calle, una provincia, un número de casa, etc.

Los contactos telemáticos se representan con objetos **`Telecom`**, que contienen el valor de la dirección como una URL, el uso o usos que tiene y el periodo de validez del mismo.

La representación de un idioma, que puede usarse para comunicarse con una persona, se hace con la clase **`LanguageCommunication.`**

La clase **`Person`** se extiende, añadiendo más semántica, para representar a un paciente. Para representar al personal sanitario se utilizaría una persona en el rol de **`Healthcare Profesional`** que también especifica la norma, como se indica en la figura 16.

5.2.3. PID

El servicio de identificación de personas (PID), especificado por el grupo CORBAmed del OMG, además de permitir el acceso a los datos demográficos de forma normalizada también provee de una serie de interfaces que estandarizan la gestión del identificador de persona. El modelo de datos, que se presenta a continuación, da soporte a estas dos tareas fundamentales.

Se definen estructuras de datos para almacenar identificadores y características demográficas. Como este servicio se define pensando en la implementación en CORBA, tanto los tipos de datos como las interfaces están descritos en IDL, aunque para facilitar la representación en esta Tesis se utilizarán diagramas de clase UML (Unified Modelling Language) [135]. Este modelo de referencia es utilizado para especificar la interfaz del servicio PID, que se verá en el siguiente apartado de este capítulo.

Identificadores

El modelo de referencia del servicio PID provee de la semántica necesaria en los procesos de identificación. Un concepto básico en el modelo del servicio PID es el dominio del identificador (ID Domain). Dentro de un dominio se mantiene un identificador único para cada identidad de persona, un IP representado en este caso por el tipo **`PersonId`**, y nunca se asignará el mismo identificador a dos personas distintas.

Idealmente hay sólo un IP para cada persona, aunque en realidad podría haber duplicados si a una persona se le asigna uno adicional dentro del mismo dominio. La gestión del identificador es local. Para utilizar el identificador fuera del dominio es necesario indicar el nombre del dominio al que pertenece ese identificador, para lo que se puede utilizar un objeto del tipo **`QualifiedPersonId`**, que precisamente incluye tanto el identificador local como el nombre del dominio, como se puede apreciar en el modelo de datos para identificadores, presentado en la figura 17.

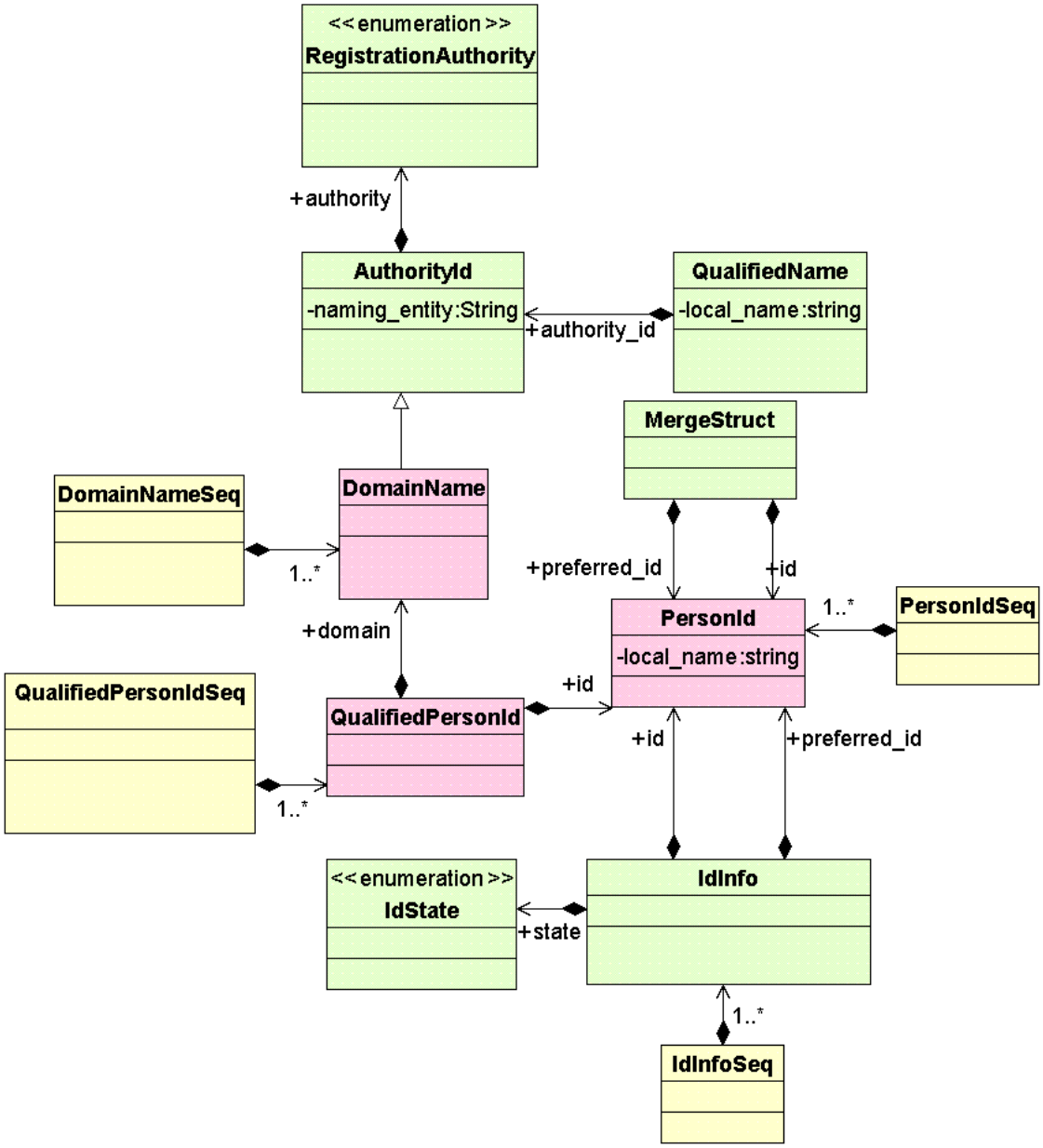


Figura 17. Modelo de información para identificadores en el PID de CORBAmed.

Aunque el tipo **`PersonId`** es simplemente una cadena de caracteres, éste se extiende para crear tipos más complejos como **`IdInfo`**, que añade un atributo para indicar el estado en el que se encuentra el identificador.

Datos demográficos

En cuanto a los datos demográficos, como se aprecia en el modelo representado en la figura 18, para dar cabida a cualquier atributo demográfico se ha creado el tipo **`Trait`**, que contiene un campo de nombre, de tipo **`TraitName`**, y otro de valor, de cualquier tipo (**`Any`** según el OMG). El modelo no especifica nada sobre los nombres de atributos,

de manera que no especifica una semántica de atributos demográficos, sólo establece los contenedores necesarios para manejarlos. La clase **`TraitSpec`** permite especificar algunas características de un atributo en concreto: si se puede utilizar en la búsqueda de personas, si es obligatorio o si es de sólo lectura.

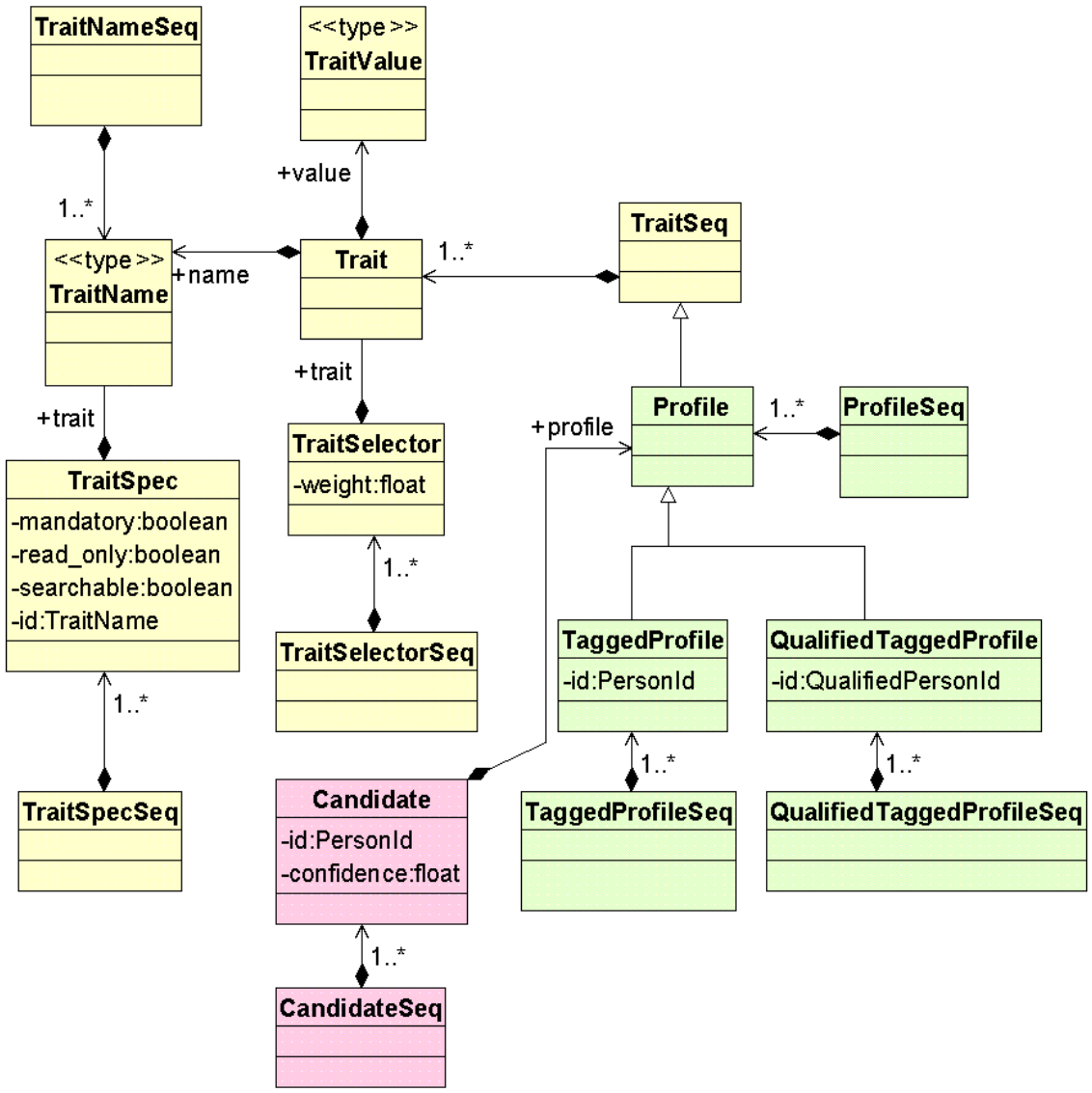


Figura 18. Modelo de información demográfica en el PID de CORBAmed

Otro concepto importante en el modelo es el de perfil o **`Profile`**, que es una secuencia de atributos. **`TaggedProfile`** y **`QualifiedTaggedProfile`** añaden al perfil el identificador local de la persona a la que pertenece o éste más el nombre del dominio respectivamente. Cuando el cliente del servicio pretende hacer la búsqueda de personas que cumplan un determinado perfil, el servidor le devuelve elementos del tipo

`Candidate`, en los que se añade al perfil el identificador de la persona a la que pertenece y el campo **`confidence`**, cuyo valor está entre cero y uno e indica la certeza que el servicio tiene de que esa persona sea la que el cliente está buscando. De este modo, cuanto mayor es la confianza mayor certeza hay en que ese candidato es el buscado: 1 indicaría certeza absoluta, porque los atributos de búsqueda coinciden plenamente con los del candidato. La norma no especifica cómo asignar y cómo utilizar este valor de confianza; estas tareas son dependientes de la implementación. También se especifica en la norma que no son comparables los valores del campo de confianza entregados por distintos servicios o en distintas llamadas al mismo servicio, ya que su significado sólo es relativo a una determinada invocación. Como se verá en las aportaciones de esta Tesis Doctoral el uso de este campo de confianza va a ser muy importante en el desarrollo de nuestra tarea.

5.3. Servicio de identificación de personas en CORBAmed

Se han visto ya los principales modelos de datos que pueden facilitar el establecimiento de una semántica común para los conceptos más relevantes manejados por servidores demográficos. Queda por analizar cómo debe ser la interfaz con estos servidores, ya que para conseguir una arquitectura que permita la integración no sólo es necesario estar de acuerdo en la semántica de la información sino también en las interfaces de acceso.

Uno de los servicios de identificación de personas más extendidos en el dominio sanitario es, como ya se ha expuesto, el dado por el grupo de trabajo CORBAmed del OMG [29]. Este servidor es el que se va a utilizar como punto de partida en esta Tesis Doctoral. A continuación pasamos a describirlo muy brevemente.

5.3.1. Conceptos básicos

Los principales objetivos del servicio de identificación de personas (PIDS, Person Identification Service) del grupo CORBAmed son:

- Establecer mecanismos para la asignación y gestión de identificadores de persona en un dominio particular.
- Facilitar la correlación de identificadores entre distintos dominios.
- Facilitar la búsqueda y localización de personas, de forma independiente a los algoritmos de localización.

- Facilitar la federación de servicios PID independientemente de la topología utilizada.
- Permitir la confidencialidad de datos personales utilizando las políticas y mecanismos de seguridad necesarios.
- Permitir la interoperatividad de servidores PID que implementen el núcleo del servicio y que puedan extenderse y personalizarse para las necesidades de cada implementación.
- Definir distintos niveles de complejidad, desde sólo búsquedas en un dominio hasta el funcionamiento en grandes federaciones con correlación de identificadores de dominio.

Para la validación de los métodos y modelos propuestos en esta Tesis se ha escogido precisamente la integración de servidores PID entre ellos y con el resto de componentes involucrados en la organización sanitaria. De manera que a continuación se presentan principalmente los aspectos del servicio relacionados con estas tareas, describiendo muy brevemente algunas de las demás funcionalidades.

5.3.2. Interfaces

El servicio se estructura en una serie de interfaces, mostradas en la figura 19, de las cuales vamos a hacer especial hincapié en tres:

- **`IdentifyPerson`**
- **`ProfileAccess`**
- **`CorrelationMgr`**

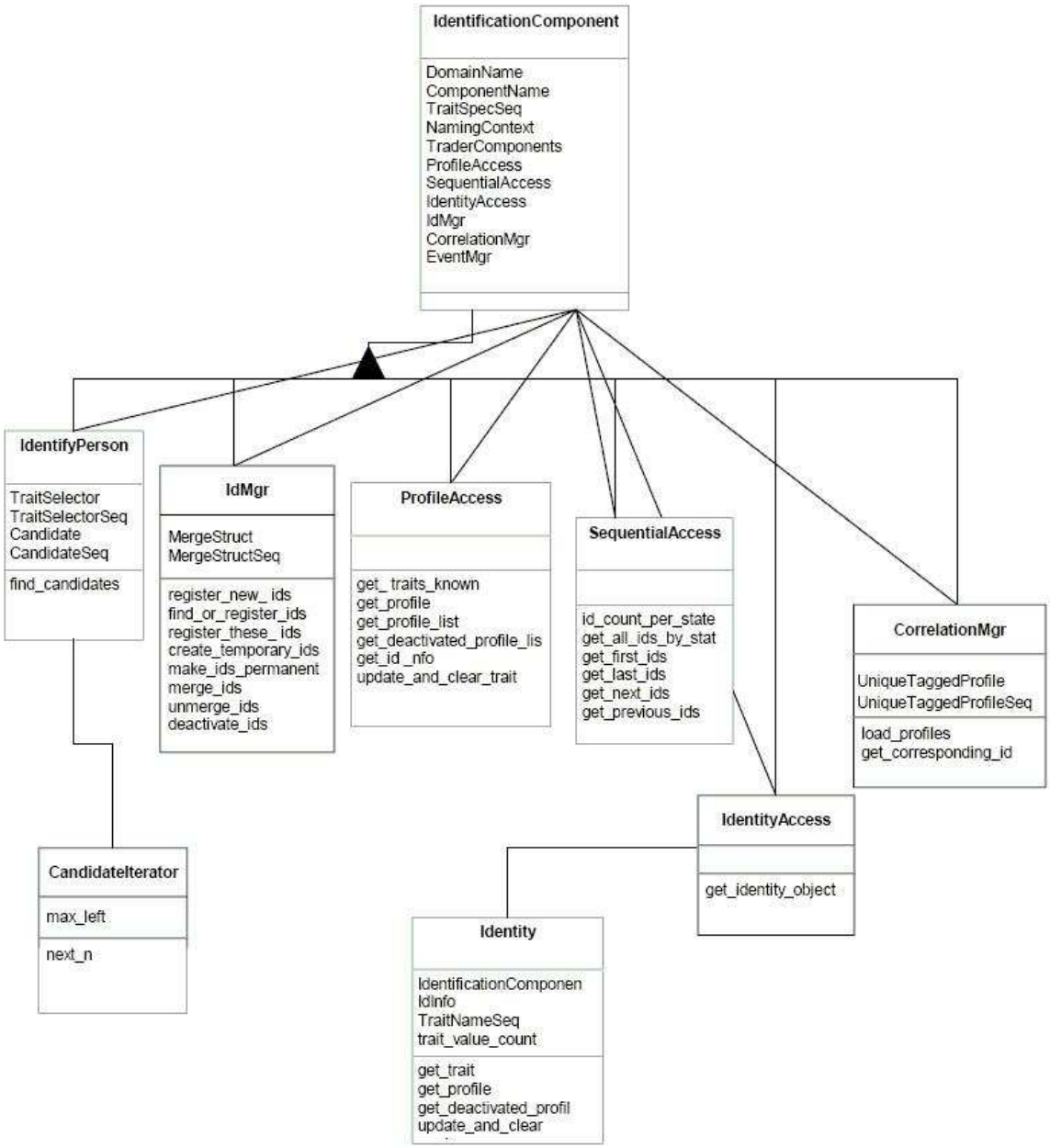


Figura 19. Interfaces del Servicio PID

Interfaz IdentifyPerson

Esta interfaz tiene un único método, **`find_candidates`,** que debe ser invocado por el cliente cuando quiera hacer una búsqueda de personas que cumplan con determinado perfil, o conjunto de atributos. El servidor devolvería una lista de candidatos cuyos atributos coinciden, total o parcialmente, con los introducidos. El grado de coincidencia de cada candidato con el perfil de búsqueda se especifica con el parámetro de confianza. Para cada candidato se devuelven también los atributos que el cliente quiere conocer

acerca de los mismos y que han sido indicados en la llamada. Los parámetros de entrada a este método son:

- **`Profile_selector`:** conjunto de atributos conocidos por el cliente acerca de la persona/s buscada/s. Es una secuencia de elementos del tipo **`TraitSelector`**. Un **`TraitSelector`** es un parámetro usado para indicar al servicio un atributo conocido y que puede utilizar para buscar una persona. Está formado por el atributo (con nombre y valor) y un campo de peso. El campo de peso sirve para que el cliente especifique la preferencia que piensa que el servidor debería dar a este atributo del perfil de búsqueda para determinar la confianza en cada candidato. Podría interpretarse como la confianza que el cliente tiene sobre cada atributo del perfil usado para la búsqueda. Debe estar entre cero y uno y son medidas relativas. Una coincidencia exacta en un atributo con peso 0,5 incrementa la confianza el doble que la coincidencia de un atributo con peso 0,25. Los algoritmos de búsqueda no están estandarizados y por tanto el uso que se haga de este campo es completamente dependiente de la implementación.
- **`States_of_interest`:** es una secuencia de estados de identificadores con la que el cliente especifica en qué estado deben estar los identificadores en los que está interesado.
- **`Confidence_threshold`:** se utiliza para limitar el número de candidatos entregados eliminando aquellos sobre los que el servidor tiene una confianza de acierto por debajo del valor especificado en este parámetro de entrada. El parámetro de confianza en el candidato, como se indicó anteriormente, mide el grado de afinidad de los atributos de búsqueda con los del candidato encontrado y la certeza de que ése sea el candidato buscado. Al no estar normalizada la asignación y uso del parámetro de confianza es completamente dependiente de la implementación.
- **`Sequence_max`:** el cliente indica el máximo número de candidatos que quiere recibir como una secuencia de candidatos. Si hay más candidatos de los indicados se devuelven en un iterador.
- **`Iterator_max`:** indica el máximo número de candidatos que quiere recibir el cliente almacenados en el iterador. Ya que el uso de iteradores puede suponer

una carga excesiva en la memoria del servidor, su uso es opcional. Si un servidor no utiliza iteradores entonces se devolverá una referencia nula a un objeto, en lugar del iterador. Los candidatos que cumplan el perfil de búsqueda y no tengan cabida en la secuencia de candidatos o en el iterador no serán entregados al cliente.

- **`Traits_requested`:** indica los atributos que el cliente quiere conocer acerca de las personas que cumplan el perfil de búsqueda. Puede indicarse que no se buscan atributos, sólo los identificadores de persona, que se quieren recuperar todos los atributos conocidos, o indicar una lista con los atributos deseados. Por lo tanto esta interfaz no es sólo de búsqueda de personas sino que también puede ser utilizado para recuperar información de las mismas.

Y los de salida serán:

- **`CandidateSeq`:** es la secuencia de candidatos entregados por el servidor después de la búsqueda. Cada candidato incluye el identificador de persona, el campo confianza que indica cómo de bien esa persona se ajusta al perfil de búsqueda, y el conjunto de atributos solicitados en la llamada.
- **`CandidateIterator`:** almacena los candidatos que cumplen el perfil pero que no caben en la secuencia de candidatos.

Interfaz ProfileAccess

Esta interfaz contiene los métodos elementales de un servicio de identificación de personas. El más significativo de estos métodos es **`get_profile`**. Con esta operación se pide al servidor un conjunto determinado de atributos de una persona. Tiene como parámetros de entrada el identificador de persona y el conjunto de atributos que se quieren conocer acerca de la misma. El método **`get_traits_known`** devuelve un conjunto de nombres de atributos. Estos nombres se corresponden con los atributos que el servidor conoce acerca de determinada persona, cuyo identificador se pasa en la llamada a este método. El método **`update_and_clear_traits`** es el único, de los vistos hasta ahora, en el que además de consultar la base de datos ésta se modifica. Se pasa como parámetro una secuencia de elementos del tipo **`ProfileUpdate`**. Este tipo contiene un identificador de persona y un perfil. De este modo, al invocar este

procedimiento se modifica el perfil de la persona a la que corresponde el identificador con los nuevos valores introducidos.

Interfaz CorrelationMgr

Esta es la interfaz a través de la cual el servicio aporta algunas herramientas para facilitar la federación de servidores PID, por ello requiere una atención especial cuando se está tratando la problemática de la integración.

Esta interfaz utiliza un atributo especial denominado **`PersonIdTraits::CorrelatedIds`,** donde se guardan las referencias a todos los IDs que están correlacionados dentro de un dominio de correlación. Esta información debe ser consistente con la obtenida al utilizar la operación **`get_corresponding_ids`,** que se describirá a continuación. Se utiliza también el atributo **`source_domains`** que almacena una lista de identificadores de dominio desde los que se pueden recibir perfiles o en los que se puede requerir correspondencias de identificadores.

Las operaciones permitidas en esta interfaz son:

- **`Load_profiles`**: tiene un único parámetro de entrada del tipo **`QualifiedTaggedProfileSeq`**, que incluye un identificador de persona, el identificador del dominio y un perfil o conjunto de atributos correspondientes a esa persona. Sirve para cargar este perfil en el servidor desde un determinado dominio de identificación. La correlación con perfiles previamente cargados desde otro dominio puede hacerse inmediatamente o no, según la implementación. Tampoco está normalizado cómo se hace esta correlación.
- **`Get_corresponding_ids`:** recibe como parámetros de entrada un **`PersonId`** más el identificador del dominio al que pertenece, es decir un **`QualifiedPersonId`**, y un conjunto de nombres de dominio (**`DomainNameSeq`**). Devolverá un conjunto de **`QualifiedPersonId`** con los identificadores, para cada dominio indicado, correspondientes al pasado en la llamada.

Se destaca aquí cómo el mecanismo para correlacionar varios dominios de identificación es cargar en el servidor los perfiles correspondientes a cada persona en cada uno de los dominios más el identificador de la persona y del dominio. Es decir que

podría verse como un mecanismo de centralización de la información de personas en un único servidor.

Otras interfaces

Todas las interfaces del PIDS heredan de la clase **`IdentificationComponent`**. En esta interfaz se guardarán referencias a cada una de las interfaces implementadas en el servidor, así un cliente puede obtener la referencia a cualquiera de los objetos que actúen como interfaces a partir del objeto **`IdentificationComponent`**.

La interfaz **`SequentialAccess`,** aporta mecanismos para moverse dentro de un conjunto de IDs. Es útil sobre todo cuando el número de identificadores es fácil de manejar por una persona.

En cuanto a la interfaz **`IdentityAccess`** es similar a **`ProfileAccess`** salvo porque se obtiene un objeto distinto para cada identificador. Esto principalmente facilita algunos mecanismos de seguridad.

Como se había indicado anteriormente el servicio PID también proporciona mecanismos para la gestión de identificadores. Éstos son utilizados a través de la interfaz **`idMgr`.**

5.4. Conclusiones

Para la incorporación a una arquitectura sanitaria, según el modelo de federación, de componentes que fueron desarrollados para realizar las tareas de identificación de personas y recuperación de información demográfica para dominios específicos, será necesario asegurar que se mantendrá la autonomía de gestión de identificadores y los mecanismos para la localización de personas y el acceso a su información demográfica en cada dominio. Al mismo tiempo estas capacidades estarán disponibles para su uso por otros componentes de la arquitectura.

En particular en las tareas de gestión de HCE, cuando ésta se compone de información almacenada en distintos dominios, el primer requisito es la correcta identificación del paciente en cada uno de los sistemas implicados y posteriormente la recuperación de su información demográfica [60,136]. Estas tareas deberían presentarse al cliente como un único servicio, de manera que se hiciera transparente todo el proceso subyacente de descomposición en dominios. Éste sería el servidor demográfico federado, identificado en varios trabajos [32,137,138] como PMI (Person Master Index), o UPI (Unique

Person Index), que por supuesto se apoyaría en el uso de los servidores demográficos de cada dominio. Si éstos han sido correctamente diseñados, previendo la posible incorporación a un dominio más amplio, se facilita la interoperatividad y se reduce el número de agentes de integración a implementar.

Un diseño basado en las normas aquí presentadas es una excelente elección. Estas normas son en parte complementarias y no totalmente excluyentes ya que no cubren exactamente las mismas necesidades de integración.

El servicio de identificación de personas ayuda a la interoperatividad funcional, ya que define una interfaz de acceso a información demográfica y mecanismos normalizados de búsqueda de personas. Sin embargo tiene una enorme lacra en cuanto a la semántica de los atributos, ya que no está definida y cada sistema, aunque se diseñe conforme a esta norma, puede utilizar distintos atributos lo que dificulta la integración semántica. Los GPIC no clínicos del CEN cubren estas necesidades semánticas ofreciendo modelos de datos demográficos e identificadores mucho más concretos y específicos. Para terminar, el openEHR ofrece un marco de trabajo más amplio integrando información demográfica, clínica y de servicios en una única arquitectura basada en dos modelos: de referencia y de arquetipos. El primero ofrece contenedores para los datos demográficos y los identificadores y el segundo un mecanismo para especificar semántica más específica, que podría estar basada en los GPIC del CEN, de forma cercana al usuario abstrayendo esta especificación de la implementación técnica.

Parte de este trabajo de Tesis se centra en la definición de arquetipos demográficos basados en los GPIC del CEN y en la especificación de un modelo de integración de servidores demográficos que cumplan con el servicio PID de CORBAmed o que sean fácilmente transformables al mismo, consiguiendo así un servicio demográfico para toda la federación. Algunos de los puntos débiles de los modelos existentes sobre los que será necesario investigar para lograr estos propósitos son:

- Será necesario incorporar a HISA mecanismos para la gestión de arquetipos y definir arquetipos demográficos que integren toda la semántica de los GPIC resolviendo carencias o posibles conflictos entre los modelos de datos de HISA y GPIC. También se tomará en consideración el modelo demográfico del openEHR, diseñado específicamente para su uso en la definición de arquetipos.

- El servicio PID no distingue tipos de personas pero en la práctica es necesario distinguirlos. El concepto de rol es utilizado en el modelo openEHR y en los GPIC para hacer esta distinción. Aparte del uso de roles en los GPIC se incluyen clases específicas para pacientes, cosa que no ocurre en el modelo del openEHR. En el servidor demográfico el rol será tratado como un atributo más siendo el cliente el que tiene que utilizar adecuadamente esta semántica. Extensiones del servidor podrían facilitar al cliente esta tarea de gestión de roles.
- Otra consideración importante es que el modelo de correlación propuesto para el servidor PID es centralizado, lo que le confiere poca escalabilidad. Es necesario cambiar este modelo hacia uno distribuido que permita la correlación de forma más dinámica. Aunque, como se verá posteriormente, la idea de correlación está basada en la interfaz **`CorrelationMgr`**, se presentará una arquitectura distribuida que resuelve conflictos de integración que no afrontaba esta interfaz, como la escalabilidad o la inconsistencia semántica.

Estos son sólo algunos ejemplos de decisiones de diseño que es necesario realizar y que se presentan en los próximos capítulos de esta Tesis Doctoral.

Capítulo 6. Resultados: extensiones a HISA

6.1. Introducción

En este capítulo veremos las principales aportaciones de esta Tesis Doctoral, que están centradas en el diseño de una metodología para la integración de sistemas de información sanitarios siguiendo el modelo de federación. Se analiza especialmente el papel que los sistemas de historia clínica deben jugar en esta arquitectura de sistemas. Las aportaciones se realizan en el marco de las directrices marcadas por la norma europea HISA: la especificación de una arquitectura abierta, basada en el estándar ODP y por tanto en la idea de computación distribuida, para la integración e interoperatividad de sistemas sanitarios.

La investigación desarrollada se va a estructurar en dos partes:

- **Resultados**, presentados en el presente capítulo, donde se muestra el trabajo desarrollado sobre la arquitectura sanitaria, detallando los requisitos de la misma, en especial los referidos a la integración, y la metodología propuesta para cumplirlos.
- **Validación de la metodología propuesta**, presentada en el siguiente capítulo, donde se abordan dos de las principales tareas en toda arquitectura de sistemas sanitarios: la identificación de personas y la gestión de información demográfica.

Aunque no se profundizará en los elementos que conforman la arquitectura se construyen los cimientos que permitan la posterior especificación detallada de cada uno de ellos. Esta idea de diseño, desde lo más general a lo más particular, es el modelo propuesto para el desarrollo de la arquitectura HISA. Se partirá de la visión general y las vistas ya especificadas en HISA, matizando y ampliando cada una de ellas. Por otro lado se realizará una revisión, en un alto nivel de abstracción, de las vistas de ingeniería y tecnológica. Este trabajo de Tesis pretende que las contribuciones realizadas sean lo más independientes posible de las tecnologías subyacentes; sin embargo se considera también una aportación interesante el análisis realizado sobre los beneficios que las nuevas técnicas de gestión semántica y computación distribuida pueden introducir. Así,

las consideraciones en las vistas de empresa, información y computacional serán aplicables a cualquier tecnología, mientras que las centradas en las vistas de ingeniería y tecnológica están basadas en las técnicas actualmente disponibles y exploran las posibilidades que éstas abren en la resolución del problema que nos ocupa.

El trabajo aquí desarrollado parte de la especificación HISA, añadiendo algunos conceptos que detallan más la funcionalidad y prestaciones esperadas de la arquitectura y los componentes comunes que se deben encontrar en ella.

Como especifica la norma es necesario comenzar presentando un documento que, de forma concisa y orientada a la gestión, identifique con un alto grado de abstracción los requisitos y los objetivos estratégicos del sistema. Esto es lo que se denomina paradigma estratégico. A continuación se deberán desarrollar las vistas independientes de la tecnología: de empresa, de información y computacional.

6.2. Paradigma Estratégico

En la revisión realizada al paradigma estratégico de HISA se añaden requisitos y nociones importantes de ODP que no están reflejadas en el original y que se consideran fundamentales para el diseño de la arquitectura de sistemas sanitarios. Por otro lado se incluyen algunas consideraciones de integración para hacer hincapié en la importancia del modelo de federación de sistemas. Para terminar se señala el papel que los sistemas de gestión de historia clínica juegan dentro de la arquitectura y se indican algunas ideas de modelado.

6.2.1. Requisitos generales

En una organización sanitaria que incluya múltiples centros, unidades e individuos, diferentes tipos de actores necesitarán compartir información y colaborar de acuerdo a determinados procesos y flujos de trabajo. Una arquitectura de sistemas de información sanitarios debe ser el soporte que permita realizar estas tareas. La información y las reglas de negocio comunes a distintos sectores de la organización estarán integradas en una capa específica de la arquitectura y deben ser accesibles a través de agentes software que implementen interfaces públicas y estables.

Los distintos sistemas implicados en la organización se añadirán a la arquitectura siguiendo un modelo federado que permita el funcionamiento autónomo de los mismos y la colaboración entre ellos para lograr objetivos específicos dentro de la federación.

De manera que otro de los requisitos de la arquitectura será la resolución de los problemas de integración a nivel semántico, funcional y de instancia.

El conjunto de sistemas que conforman la federación no es estático por lo que pueden incluirse, eliminarse o modificarse sistemas. La arquitectura deberá ser escalable y facilitar las modificaciones locales evitando la interrupción de los servicios generales proporcionados.

Aunque los servicios finales ofrecidos por una organización sanitaria están relacionados con la atención al paciente dentro de la organización, es necesario también realizar otras tareas, por ejemplo de investigación, administrativas o de gestión. De manera que en la labor asistencial se gestiona información clínica, pero hay otros muchos factores y tareas involucrados, manejándose también información financiera, de seguridad, de identificación y de naturalezas muy diversas. Es requisito fundamental que en la arquitectura se contemplen estas necesidades y se incluyan los mecanismos necesarios para cubrirlas.

En la asistencia al paciente hay implicados numerosos profesionales sanitarios. Cada uno de éstos profesionales realiza tareas específicas de manera que necesita consultar una información concreta acerca de un paciente determinado y tras su intervención se genera nueva información que también es almacenada. En este proceso se puede hablar de gestión de conocimiento y podría dividirse en tres fases.

1) Por un lado la vista de la información acerca del paciente que el profesional consulta en un instante dado está particularizada para un contexto que depende de factores como:

- La patología y estado de la misma en el paciente.
- Factores personales del paciente (clínicos y no clínicos).
- Tipo de asistencia (urgencias, consulta externa, planta…).
- Profesional que atiende al paciente (servicio, cargo, modo de trabajo personal…).

Al introducir estos parámetros en la recuperación de la información almacenada se está personalizando la misma y se puede hablar de una extracción de conocimiento. La arquitectura debe incluir componentes que faciliten la gestión de contextos y la extracción de conocimiento.

2) A continuación el profesional estudia el conocimiento extraído, realiza una serie de procedimientos (anamnesis, exploración, pruebas de laboratorio…) y aporta su propio conocimiento al caso clínico que le ocupa. Se podría hablar de una generación de conocimiento. Por supuesto esta tarea no es fácil e implica un esfuerzo considerable del personal sanitario, que necesita consultar los procedimientos realizados, otros casos clínicos similares,etc. La arquitectura puede incluir prestaciones que ayuden al profesional en esta generación de conocimiento, no sólo facilitándole la consulta de la información necesaria sino también con herramientas que la gestionen para generar nuevo conocimiento, utilizando técnicas activas como, por ejemplo, el modelado.

3) Para terminar es necesario almacenar el conocimiento generado para que sea utilizado en posteriores contactos con el paciente. La arquitectura debe incluir componentes y técnicas de representación que permitan un almacenado fiel de las conclusiones extraídas así como de los procedimientos realizados al paciente, y que facilite la explotación de los mismos.

La información implicada en la gestión de conocimiento clínico puede encontrarse en sistemas externos a la federación. Por ejemplo la medicina del futuro manejará información genética y proteómica que pueden encontrarse en sistemas que no pertenecen a la arquitectura de la organización sanitaria [12,13,132,139]. Por otro lado sistemas externos podrían querer, y tener permiso para, extraer datos de la arquitectura, así estudios epidemiológicos o sistemas de educación para profesionales sanitarios pueden verse beneficiados del conocimiento almacenado en el conjunto de la organización. Es imprescindible por tanto que la arquitectura incluya elementos que permitan la comunicación y colaboración entre sistemas. Esta comunicación puede estar basada en el intercambio de mensajes según determinado estándar o considerar las posibilidades de integrar estos sistemas externos a la arquitectura de una forma mucho más acoplada.

Analizando estas premisas, la arquitectura de integración debe cumplir al menos los siguientes requisitos:

- Deberá poder integrar y acomodar todos los datos y reglas de negocio relevantes a las actividades clínicas, administrativas y de gestión en la organización sanitaria.

- Será abierta en todos los niveles y permitirá extensiones y evoluciones según los requisitos locales, leyes nacionales y evolución de los estándares.
- Representará un conjunto autónomo y reemplazable de componentes de la arquitectura, separado de otros usos. El sistema completo se verá como una federación de sistemas que interactuarán para conseguir objetivos comunes.
- Existirán componentes que faciliten la comunicación con sistemas externos a la arquitectura.
- Soportará al menos los requisitos identificados en la vista de empresa del estándar HISA y las extensiones propuestas en esta Tesis Doctoral, facilitando además el crecimiento de estos requisitos.
- Implementará al menos el modelo de información especificado en la vista de información de HISA y las extensiones propuestas en esta Tesis Doctoral. Este modelo de información es la base de la integridad semántica del sistema y representa la ontología base y común que todos los sistemas de la arquitectura deben manejar. Aunque el modelo de información es propio de la vista de información, se introducirán elementos que faciliten la gestión y ampliación de la ontología en las restantes vistas.
- Proporcionará al menos las prestaciones especificadas en la vista computacional del estándar HISA y las extensiones propuestas en esta Tesis Doctoral. Esta especificación es la base de la integridad funcional. Resolver este problema en la vista computacional y no en las inferiores, de tecnología e ingeniería, asegura una solución independiente de implementaciones concretas.
- Todas las vistas deberán incluir elementos para solventar los problemas de integridad de instancia. Es importante identificar la información altamente duplicada, como la demográfica, para prestar especial atención a los servicios que gestionen esta información ya que los problemas de integridad en el nivel de entidad serán mucho más frecuentes en estos casos. La arquitectura debe asegurar:
    - La correcta identificación de entidades reales: resolver qué objetos corresponden a la misma entidad real.

- Solucionar los conflictos en los valores de los atributos de las entidades reales: objetos correspondientes a la misma entidad pueden tener valores en los atributos que no concuerden. Como los sistemas se añaden a la arquitectura con un modelo federado, van a seguir funcionando de forma autónoma, y aunque se resuelvan los problemas en los niveles semántico y funcional los problemas a nivel de entidad pueden ser frecuentes.

Como conclusión a esta visión general de requisitos de la arquitectura, señalar que no se está hablando aquí de un sistema de historia clínica que permita la integración de sistemas heterogéneos, tarea que no es nada fácil, sino de una arquitectura de integración que permita la colaboración de agentes para lograr proporcionar a los usuarios de una organización sanitaria las prestaciones que necesitan en su actividad diaria. Esto implica, además de la integración de la información de historia clínica, la integración de otra información, incluso procedente de sistemas externos, que va a necesitarse para el desarrollo de aplicaciones que cubran las necesidades en las actividades del usuario de la arquitectura, y en particular aquellas relacionadas con la extracción, asistencia a la generación y almacenado de conocimiento.

6.2.2. Consideraciones generales de arquitectura. Extensión de los términos HISA.

La arquitectura está fundamentada en el modelo ODP, y por tanto está basada en la descomposición de las tareas a realizar dentro de la organización sanitaria y la asignación de estas tareas a componentes o agentes software. En esta Tesis Doctoral se utilizará el término componente, agente y objeto indistintamente. Para aclarar esta decisión y no dejar lugar a dudas o posibles confusiones, se entiende que un componente, agente u objeto será un elemento computacional que se podrá relacionar con otros a través de interfaces. Estos elementos serán reactivos o proactivos, es decir que responderán a llamadas y eventos externos o generarán llamadas a otros componentes cuando, de forma autónoma, lo consideren necesario. Tal y como indica ODP los agentes sólo pueden interactuar a través de interfaces.

La arquitectura HISA está basada en la descomposición en servicios y por tanto las interacciones propuestas son básicamente operaciones. En la extensión de la arquitectura aquí presentada se amplía esta idea para incluir la posibilidad de utilizar flujos y señales.

En la visión general de la arquitectura es importante señalar un aspecto que, en principio, la arquitectura HISA no deja del todo claro. HISA define la vista computacional en términos de servicios básicos y compuestos; es necesario aclarar el término servicio para que no resulte confuso o restrictivo.

Tal y como se indicó en el capítulo dedicado a ODP el concepto Servicio se entiende como **"el conjunto de capacidades proporcionado por un objeto en una interfaz computacional"**. Con esto se pretende clarificar que dentro de la arquitectura HISA existe la posibilidad de establecer vínculos entre agentes que no sigan estrictamente el paradigma cliente/servidor, con uno de ellos actuando siempre como invocador de métodos ofrecidos por el otro. Las formas de colaboración entre agentes pueden ser muy diversas y el papel jugado por un agente u objeto puede diferir en las distintas interacciones con otros objetos.

El concepto servicio utilizado en HISA se interpreta por tanto en esta Tesis como el conjunto de prestaciones ofrecidas por un agente a través de sus interfaces, en las que el objeto no tiene que asumir necesariamente el papel de servidor. Estas prestaciones pueden ser llevadas a cabo por composición, gracias a la interacción del agente con otros a través de interfaces en las que el tipo de interacción y el papel del objeto pueden ser cualquiera. Esta idea será especialmente importante al especificar la vista computacional.

6.2.3. Consideraciones de integración

La definición formal de interfaces en términos de firma, comportamiento y contrato de entorno es la base de la integridad funcional y se realiza principalmente en la vista computacional.

La especificación de ontologías comunes a los agentes de la arquitectura, realizada en la vista de información, y de mecanismos para gestionar adecuadamente estas ontologías asegura la integridad semántica.

La integridad de instancia se debe asegurar en todos los agentes que gestionen información que provenga de distintos dominios y cuando ésta pueda representar a la misma entidad real. Especialmente en la vista de información será necesario señalar qué grado de duplicidad puede haber en determinado tipo de información para que esto sea considerado en las demás vistas a la hora de resolver conflictos de entidad. Así, por ejemplo, la vista de información del sistema señalará que cierta información

demográfica de un paciente es una información altamente duplicada, mientras que los datos de enfermería recogidos durante un ingreso hospitalario del mismo es información poco duplicada. Esta visión remarca el hecho, ya señalado en HISA, de que las distintas vistas estarán relacionadas entre sí, de manera que las consideraciones realizadas en una de ellas pueden implicar determinadas medidas en las otras.

Otro aspecto a considerar es el hecho de que en los sistemas de información sanitarios el paradigma más habitual de integración de sistemas es el de federación, es decir una comunidad de dominios, creados de forma independiente, que van a colaborar para alcanzar objetivos comunes pero que mantendrán su autonomía. Los dominios incluidos en una federación pueden ser administrativos y/o tecnológicos.

La federación de dominios administrativos está relacionada con el interfuncionamiento entre dominios de la misma o de distintas organizaciones para proporcionar el uso compartido, la integración o la partición de recursos y aplicaciones a través de distintos sistemas y ubicaciones en respuesta a las necesidades particulares de los usuarios. Los límites administrativos constituyen también puntos en los que cambia la responsabilidad de gestión, para aspectos como la seguridad, asignación de recursos, etc. Mantener la autonomía de estos dominios tras la integración permite la evolución independiente de las políticas de gestión de cada uno de ellos.

La federación de dominios tecnológicos está relacionada con la integración de distintas arquitecturas de sistemas o de sistemas con recursos y prestaciones diferentes, proporcionando la modularidad necesaria para permitir que el crecimiento no afecte a las aplicaciones existentes. Es frecuente que coincidan los dos tipos de federación ya que las diferencias en la administración suelen conducir a la adopción de distintas soluciones tecnológicas.

6.2.4. Consideraciones sobre el papel de las aplicaciones y componentes de gestión de historia clínica dentro de la arquitectura HISA

En las actividades puramente asistenciales y orientadas a un paciente, la consulta y modificación del historial clínico del mismo es crucial, aunque sólo representa una parte del conjunto de tareas que se llevan a cabo durante estas actividades, y en las que no siempre se gestiona información estrictamente perteneciente a la historia clínica del paciente. Por otro lado, la información almacenada en sistemas que se crearon con el propósito de gestionar la historia clínica puede ser de gran interés para otras actividades

no directamente asistenciales o centradas en un paciente, como por ejemplo la investigación científica, la gestión de recursos o el análisis de costes y eficiencia de la organización.

Un diseño eficiente de un sistema de gestión de historia clínica tomará en consideración, desde el primer momento, que deberá integrarse en una arquitectura de sistemas sanitarios y que las aplicaciones que se van a desarrollar sobre la misma pueden tener requisitos muy distintos. Dada la amplia cobertura de la norma HISA, es evidente que las necesidades específicas de gestión de la historia clínica única no están suficientemente detalladas, y afrontar el diseño conforme a HISA de los componentes de la arquitectura que provean las prestaciones necesarias requiere un análisis y planificación mucho más profundos. Desde el principio la metodología de diseño del sistema de historia clínica debe considerar el paradigma de computación distribuida, es decir debe consistir en una serie de agentes con funcionalidades muy bien definidas que van a componerse para ofrecer prestaciones más complejas. Esto por un lado tiene la ventaja de poder aprovechar los componentes que cubran parte de las necesidades de este sistema de información y que ya existan en la arquitectura sanitaria, y por otro facilita la incorporación del sistema de gestión de historia clínica dentro de la federación sanitaria.

La normalización de una arquitectura específica para la gestión de la historia clínica debe ser conforme a ODP y a HISA, de manera que partiendo de las vistas de empresa, información y computación especificadas en HISA se deben realizar las extensiones necesarias e identificar cada nuevo componente como extensión de los propuestos por la norma raíz. Por supuesto será tarea de los desarrolladores profundizar en los niveles más bajos de la especificación ODP. Siguiendo este paradigma de diseño se puede hablar de una arquitectura de sistemas de historia clínica que cubren determinadas prestaciones dentro de la más amplia de sistemas sanitarios, HISA.

Al profundizar en el punto de vista de empresa será necesario identificar en qué actividades de usuario se requiere el uso de la historia clínica y cómo se va a utilizar la misma. En cuanto a la vista de información, partiendo de la especificada en HISA será necesario extenderla con información más concisa respecto al dominio de la historia clínica. Los esfuerzos de normalización realizados por el CEN en el EN13606-1 ya han proporcionado un modelo de información, de manera que sería lógico seleccionarlo como vista de información para los componentes de gestión de HCE. En este caso el

esfuerzo es de consenso entre este modelo y la vista de información de HISA. En la vista computacional el trabajo de ampliación es mucho mayor, ya que será necesario hacer una descomposición bien planificada en componentes, detallar las prestaciones e interfaces de cada uno, y definir los mecanismos de composición de los mismos.

Sin embargo, como se ha indicado anteriormente, existen por un lado varios grupos trabajando en la normalización de la historia clínica electrónica, y por otro lado numerosos sistemas de gestión de historia clínica propietarios desarrollados de espaldas a estándares y normas. De manera que es evidente que existirán sistemas ya desplegados que no se han desarrollado según el modelo funcional y de información propuesto en la arquitectura de integración, pero si ésta está correctamente diseñada debe incluir mecanismos para facilitar la incorporación de estos sistemas. Por tanto el diseño de los componentes de la arquitectura que participan en la gestión de historia clínica electrónica única debe tomar en cuenta estas consideraciones y afrontar la integración como una federación de sistemas heterogéneos.

6.3. Extensión de la vista de empresa

La arquitectura presentada en el estándar HISA está basada en un primer análisis de las actividades que los usuarios de un sistema sanitario van a ejecutar. Estas actividades, propias de la capa más alta de aplicación, se descomponen en tareas más elementales que serán realizadas por componentes especializados, que serán reutilizables en distintas actividades. Este enfoque facilita el crecimiento del sistema y mejora la escalabilidad.

En la vista de empresa especificada en HISA se han identificado tres flujos de trabajo directamente relacionados con las actividades asistenciales y algunas tareas complementarias que se refieren a actividades de gestión también imprescindibles en una organización sanitaria. Esta descomposición de tareas a alto nivel se utilizará para clasificar las actividades dentro de la organización en las siguientes categorías:

- **Flujo del sujeto de cuidado**, centrado en el paciente: actividades relacionadas con la gestión de información personal y estadística de los pacientes y con la gestión de encuentros con el paciente dentro de la organización.
- **Flujo de gestión de actividad**, centrada en la atención: actividades de gestión, control de estados y coordinación de las tareas y actividades necesarias en la organización.

- **Flujo de información clínica**, relativo a actividades de gestión de datos clínicos.
- **Gestión de usuarios y autorizaciones de acceso.**
- **Gestión de recursos** disponibles en la organización y de reglas y criterios de utilización de los mismos. Se entienden aquí por recursos los elementos necesarios para el trabajo de la organización, que pueden incluir personal, materiales, equipamiento, etc.
- **Gestión de clasificaciones, criterios de codificación y diccionarios** que proporcionan la integridad semántica.

En esta Tesis se amplia esta clasificación en los siguientes términos:

- Los requisitos de seguridad especificados en la vista de empresa se detallaban en las actividades de gestión de usuarios y autorizaciones. En esta Tesis Doctoral se extienden estos requisitos del modo siguiente:
    - Se introduce el concepto de **recurso de la arquitectura**, que se entiende como cualquier información, independientemente del formato de la misma, o funcionalidad disponible en la arquitectura de integración. Para no confundir estos recursos con los propios de la organización las actividades de gestión de recursos pasan a denominarse actividades de gestión de recursos de la organización.
    - Se distinguen más claramente las tareas de gestión de usuarios y credenciales de las tareas de gestión de recursos, que relacionan las credenciales con las autorizaciones de uso de los recursos de la arquitectura. Esta matización a HISA simplemente pretende facilitar el desarrollo de sistemas conforme a la norma al separar desde el primer momento actividades que, aunque ciertamente son muy próximas, pueden diferenciarse claramente y apoyarse en componentes con funciones muy específicas.
    - Se añaden las actividades de **auditoría**. En éstas se consideran los mecanismos para gestionar las distintas versiones de la información tratada, para registrar los accesos a los recursos de la arquitectura y para la gestión de informes referentes a este tipo de eventos. Las actividades

de usuario de esta categoría básicamente se dedican a hacer cumplir la legislación aplicable dentro de la organización sanitaria en cuanto a gestión y protección de datos. En HISA no aparecían explícitamente como tareas a realizar, sin embargo los avances legislativos en cuanto a gestión electrónica de información altamente sensible obliga a considerar, desde el mismo momento de modelado del sistema, estas tareas. Por otro lado estas actividades son suficientemente autónomas para considerarlas como independientes de las demás tareas de usuario.

- Las tareas de gestión de clasificaciones, criterios de codificación y diccionarios se amplían para contemplar el más amplio concepto de **gestión de semántica**. Se incluye no sólo la gestión de sistemas de tesauros, como en principio la norma señala, sino todas las tareas relativas a la gestión de ontologías, anotación de información, organización de modelos de información, etc. Una de las principales aportaciones de esta Tesis, como se detallará más adelante, es considerar que la vista de información no es estática y por tanto la arquitectura debe proveer de mecanismos para modificarla y utilizarla de forma dinámica. Esta gestión debe ser suficientemente sencilla para que la realicen los propios usuarios de la arquitectura sin la necesidad de la participación de los ingenieros de sistemas. Como puede apreciarse esta idea está fundamentada en el concepto de modelo de referencia y modelo de arquetipos usada en los sistemas de historia clínica. La aportación es hacer la consideración de que el problema de gestión semántica dentro de una arquitectura de sistemas sanitarios no es exclusivo de los sistemas de historia clínica. Por otro lado si el diseño de aplicaciones relacionadas con las actividades de gestión de arquetipos, y en general semántica, se fundamenta en la creación o reutilización de componentes básicos que pertenecen a la arquitectura de sistemas se tendrá un diseño mucho más óptimo, escalable e integrado.
- Se añaden las tareas de comunicación donde se tratan las actividades relativas a la conexión con sistemas que no pertenecen a la arquitectura, incluyendo, entre otros, la conversión de formatos y la definición de protocolos de comunicación con el exterior. También se incluyen actividades relativas a las comunicaciones entre componentes dentro de la propia arquitectura.

La nueva clasificación de actividades de usuario dentro de la arquitectura sanitaria y la relación con la propuesta en HISA se muestra en la figura 20.

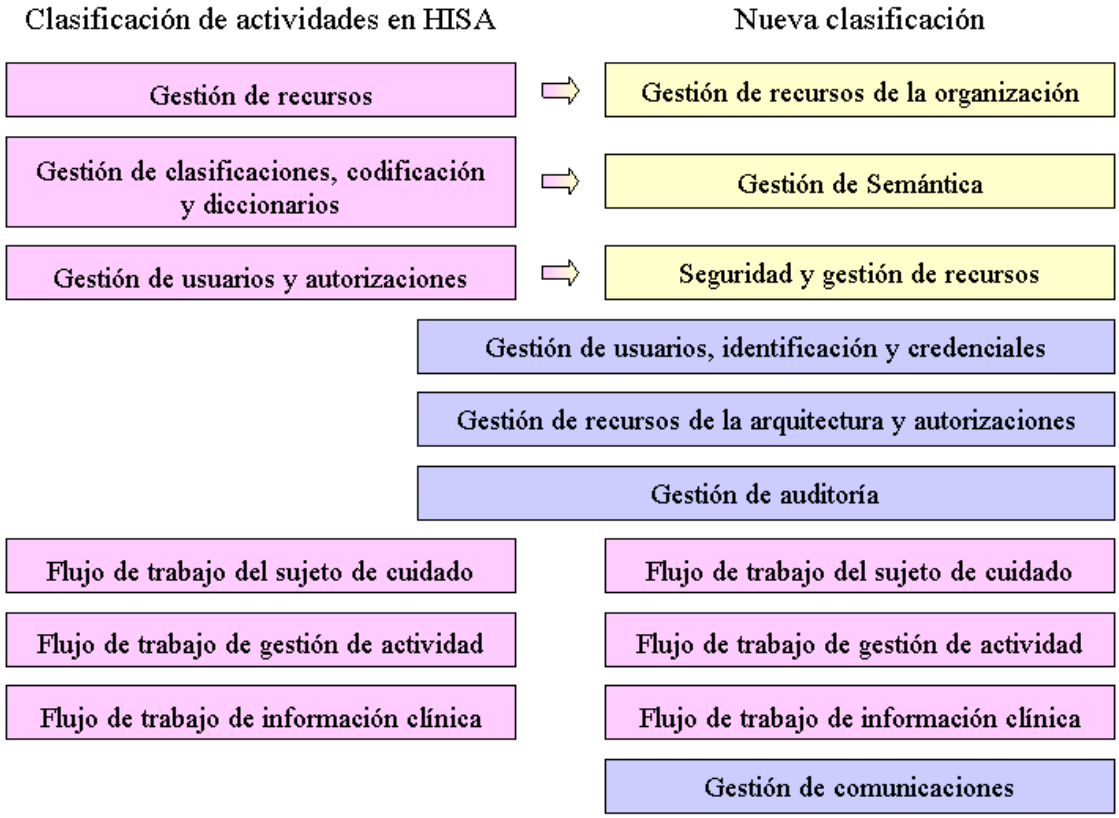


Figura 20. Modificaciones propuestas a la clasificación de actividades en HISA

Es importante también destacar que aunque se identifiquen actividades de usuario, es decir totalmente ligadas a actores del sistema y por tanto de muy alto nivel, estas actividades están relacionadas o divididas en otras más alejadas de los agentes humanos. La arquitectura debe reflejar también estas actividades aunque no sean directamente de usuario, ya que es imprescindible tenerlas en cuenta para realizar un diseño estructurado, modular, flexible y escalable.

Existen trabajos sobre la formalización de lenguajes para la representación de la vista de empresa de un sistema diseñado según ODP [140]. Aunque HISA es independiente del lenguaje de especificación que se utilice, recomendar o incluso normalizar la forma de representación de la vista de empresa podría facilitar el desarrollo de sistemas distribuidos dentro del entorno sanitario.

6.3.1. Reflejo del modelo federado en la vista de empresa

Como se ha comentado en capítulos anteriores la evolución de los sistemas sanitarios tiene repetidamente por consecuencia la fusión de subsistemas previamente existentes, gestionados independientemente, para compartir información y lograr objetivos comunes. Este modelo de colaboración se corresponde con una federación, es decir, la reunión de varios grupos que responden a autoridades diferentes (por lo que pueden representarse como dominios distintos), de modo que, juntos, puedan cooperar para alcanzar algún objetivo, pero manteniendo la autonomía de cada uno. Sin embargo la descripción actual de la vista de empresa en HISA sólo hace especial hincapié en esta circunstancia en un anexo informativo.

En este anexo se indica que en un nivel alto de abstracción, cualquier organización sanitaria puede ser descrita con un modelo federado en el que un conjunto de unidades organizativas o dominios interaccionan para ofrecer determinados servicios. Cada dominio tiene un cierto nivel de autonomía e independencia en términos de la información gestionada y las actividades soportadas. Cada sistema sólo presentará a la federación las funciones e información de interés común, sin modificar las específicas de la aplicación. Tanto los modelos de información como funcionales de los componentes de cada sistema que se vayan a integrar en la arquitectura deben ser abiertos y bien documentados para que puedan utilizarse en las aplicaciones que se desarrollen.

Sin embargo, a lo largo de la definición de las distintas vistas de la arquitectura será necesario considerar en más detalle la especificación de mecanismos para la creación de federaciones y la expresión de las reglas que deben regirlas, ya que estos términos deben constituir una parte importante de la arquitectura de sistemas sanitarios. El modelado de una federación implica la especificación de los objetivos para el interfuncionamiento entre diferentes dominios y de las políticas que rigen ese interfuncionamiento. Por ejemplo, una especificación de empresa puede incluir reglas de los dominios que especifiquen:

- Las reglas que determinan la calidad de miembro de un dominio.
- Las reglas de interacción entre dominios del mismo tipo.
- Las reglas de denominación de dominios.

Es necesario considerar que los componentes que forman parte de la arquitectura sanitaria pueden actuar sobre un único dominio o pueden interaccionar con otros para trabajar sobre un conjunto de los mismos o incluso sobre toda la federación. Esto significa que la información gestionada por un agente puede pertenecer a un sistema de información de un dominio concreto o será información perteneciente a un conjunto de sistemas de información de distintos dominios de la arquitectura. A partir de este momento, nos referiremos a **agentes/componentes autónomos** o a **agentes/componentes federados** para diferenciar estas dos posibilidades: los primeros permiten el acceso a sistemas de información dentro de un dominio; los segundos ofrecen sistemas de información federados, en el sentido de que la información gestionada pertenece a distintos dominios, y cada uno de los ellos es responsable de la información que contiene.

En referencia a los dominios administrativos, cada una de las administraciones puede desear imponer sus propios controles de acceso para fines tales como la seguridad, la contabilidad y la supervisión, además de los controles impuestos por los propios componentes. Las fronteras administrativas son también los puntos en que se producen los cambios de la responsabilidad de la gestión para asuntos tales como la atribución de recursos y las garantías de seguridad de funcionamiento.

En conclusión, al presentar la vista de empresa se extenderá la norma HISA para mostrar las características propias del modelo federado que se impone en las organizaciones sanitarias.

6.3.2. Flujo de trabajo orientado al sujeto de cuidado (paciente)

Antes de comenzar con las tareas pertenecientes a este flujo de trabajo haremos una pequeña reflexión sobre la orientación de centralización en el paciente que actualmente se realiza en la mayoría de los sistemas sanitarios. La finalidad de cualquier sistema sanitario es proveer de una serie de servicios al sujeto de cuidado o paciente para atender sus necesidades de asistencia sanitaria. En ese sentido toda la organización está centrada en el paciente. De hecho, como hemos señalado anteriormente, el documento de historia clínica se refiere a un individuo concreto, y la integración de esta información distribuida, paso crucial para una asistencia de calidad, es el objetivo de numerosos trabajos de investigación [57,59]. El tratamiento personalizado del paciente para proporcionarle la mejor atención posible es siempre el objetivo fundamental. Sin

embargo, en un análisis de una arquitectura abierta de sistemas de información clínicos federados, y aunque el servicio final de la organización es la atención al paciente, no todas las tareas están centradas en el paciente, o dicho de otro modo, los elementos de información que gestionan los componentes de la arquitectura no son siempre relativos a pacientes o a un paciente concreto.

Los agentes que se consideran dentro del flujo de trabajo orientado al sujeto de cuidado son los responsables de las tareas de identificación del individuo y de la gestión de la información personal, clínica y epidemiológica que puede ser necesaria para asegurar la consistencia de análisis estadísticos, epidemiológicos y sanitarios llevados a cabo por distintas organizaciones. También se incluyen en esta categoría las actividades de gestión de contactos del paciente con la organización sanitaria.

Una tarea esencial será la identificación unívoca del paciente dentro de la organización y la gestión adecuada de los identificadores relacionados con el mismo. Este conjunto de tareas ha sido especialmente considerado en esta Tesis ya que se ha utilizado como escenario de validación de los métodos desarrollados. Estos resultados se mostrarán más adelante; aquí sólo se hacen algunas consideraciones relativas al modelo federado que en principio no están incluidas en la norma HISA.

En el flujo de trabajo del sujeto de cuidado la relación del paciente con la organización sanitaria se define en términos de contactos o tratamientos. Un contacto o tratamiento tiene lugar para resolver una necesidad concreta de asistencia sanitaria; podría generalizarse como un proceso asistencial. Empieza con el registro del paciente, en caso de que éste no estuviera previamente registrado, y termina al final del tratamiento con la generación de los informes relativos al mismo.

La primera tarea dentro de este flujo de trabajo es el registro de la información relativa a una persona. Esta información es independiente de los futuros contactos con la organización sanitaria. El modelo de federación impone algunas reflexiones adicionales a las consideraciones realizadas en la vista de empresa de HISA. El caso más habitual es que el registro de la persona se haga dentro de un dominio de la federación, en ese caso se utilizarán los agentes autónomos encargados de estas tareas. Cuando una persona se registra en un dominio su información podrá ser consultada, siempre que se tenga la autorización necesaria, por cualquier agente del sistema. Será habitual el caso en el que la persona estuviera previamente registrada en otro dominio; esto demuestra que la información de registro de una persona está altamente duplicada en los distintos

dominios de la federación y los problemas de inconsistencia de valores serán habituales. Los procedimientos de registro en un dominio pueden tomar en consideración ésto, de modo que los agentes implicados pueden interactuar con otros para realizar registros más eficientes. Podría considerarse incluso la posibilidad de tener un registro centralizado de personas, lo que disminuiría la duplicidad de información y por tanto la necesidad de mecanismos para la resolución de conflictos de entidad. Sin embargo, esta opción puede limitar la escalabilidad del sistema y la integración del mismo en una arquitectura federada más compleja. Descentralizar el registro de pacientes implica duplicidad de información y, seguramente, de identificadores, haciendo las tareas más complejas pero, al mismo tiempo, proporcionando un sistema más escalable.

La existencia de identificadores administrativos para los sujetos de cuidado que aseguren referencias unívocas en todos los dominios es deseable, sin embargo una arquitectura escalable debe proveer mecanismos que no se basen únicamente en esta posibilidad, ya que siempre pueden existir sujetos de cuidado que no tengan asignados estos identificadores o pueden existir errores en la asignación de los mismos. Un modelo más exhaustivo será expuesto en la vista de empresa de la arquitectura de agentes demográficos que se presentará en el capítulo de validación de esta Tesis Doctoral.

Como se especificó anteriormente también pertenecen a esta categoría los componentes de la arquitectura que facilitan la gestión y seguimiento de los contactos del sujeto de cuidado con la organización. En la planificación de contactos puede considerarse información de muy diversa naturaleza: información de referencia, de planificación de recursos, de información clínica o demográfica de interés, etc.

Considerando de nuevo el modelo federado, se pueden gestionar contactos dentro un dominio si las tareas asistenciales al individuo no requieren de la intervención de elementos fuera del mismo; por ejemplo, una cita rutinaria con el médico de atención primaria. Pero se pueden gestionar contactos multidominio, si en la interacción con el paciente deben cooperar agentes pertenecientes a distintos dominios. Sería el caso de una cita para una acción quirúrgica en la que, implícitamente, se realicen citas en distintos dominios como un laboratorio externo, el servicio de anestesia o el de cirugía. Estas tareas implicarán a distintos agentes, incluso pertenecientes a otros flujos de trabajo como el de gestión de actividades. Del mismo modo se puede hacer un

seguimiento histórico de contactos del paciente dentro de un dominio o en un conjunto de dominios de la federación.

Durante un tratamiento específico la responsabilidad asistencial del paciente puede ir cambiando. Los componentes de la arquitectura que gestionan estas responsabilidades también pertenecen a este flujo de trabajo; así como todos aquellos que se encargan de gestionar el fin de un proceso asistencial y de la elaboración de los informes necesarios, ya sean clínicos o administrativos, por ejemplo para el cálculo de costes del proceso. Dado que los actores responsables de un proceso asistencial al paciente pueden pertenecer a distintos dominios podría considerarse que éste se descompone en varios procesos, cada uno relativo a uno de los dominios implicados.

La figura 21 resume algunas de las principales tareas implicadas en este flujo de trabajo y muestra la idea de federación señalando la posibilidad de que los agentes trabajen dentro de un dominio o interactúen con otros agentes para conseguir ofrecer estas prestaciones sobre distintos dominios.

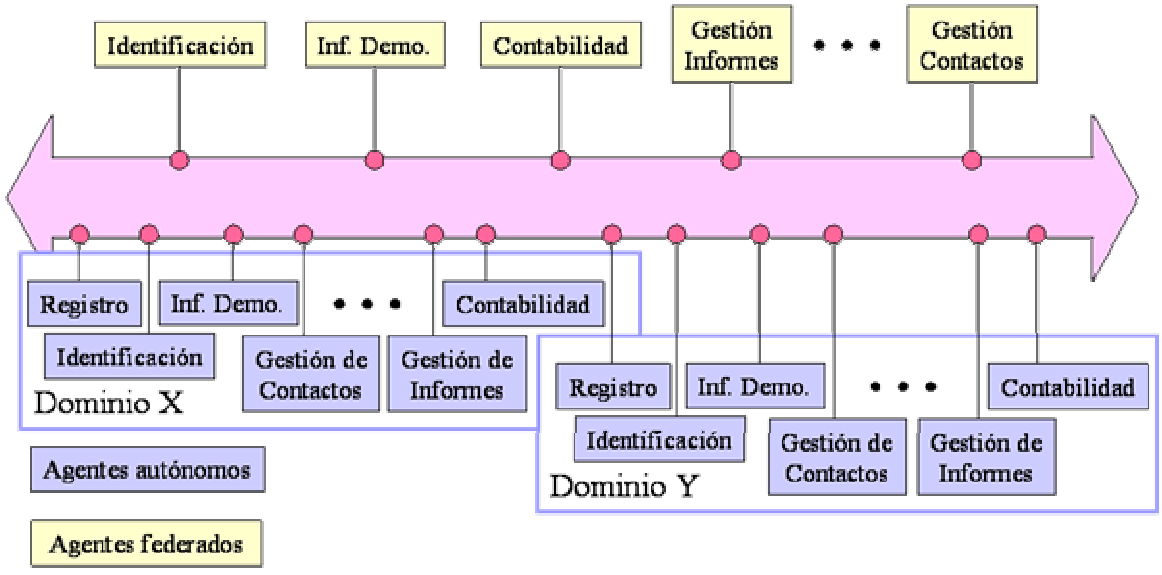


Figura 21. Principales componentes del flujo de trabajo del sujeto de cuidado.

6.3.3. Flujo de información clínica

A este flujo de trabajo pertenecen las tareas que impliquen la gestión de información clínica. Son típicas de los sistemas de laboratorio, PACS o RIS. La norma especifica que esta información puede ser elemental o la agregación de múltiples datos. Esta idea coincide con la definición de observación clínica, heredada del servicio COAS de

CORBAmed, que se utiliza para identificar cualquier registro de información que sea de carácter clínico y que puede tener cualquier naturaleza: anotaciones textuales, resultados de laboratorio, imágenes, etc. Una observación puede ser simple o puede componerse de observaciones más elementales. A partir de ahora en este trabajo se hablará de observación clínica para referirse a cualquier elemento de información clínica.

HISA señala que cada componente elemental de una observación, es decir cada observación simple, debe estar almacenada sólo una vez en el sistema para disminuir el riesgo de inconsistencias. El modelo federado puede cumplir perfectamente esta premisa, la información clínica se gestiona a través de componentes del dominio que es responsable de la misma, manteniéndose así la autonomía de cada sistema. De manera que la información clínica, al contrario que la demográfica, es información que no debe duplicarse en distintos sistemas de la federación. Sin embargo, si se analizan las tareas asistenciales sí existe duplicidad de observaciones clínicas, sobre todo cuando se realizan informes con determinados objetivos; referencias, altas, pruebas diagnósticas, etc. Estos informes pueden contener datos clínicos, extraídos de algún dominio y, por motivos fundamentalmente legales, deben mantenerse en los sistemas de información que los emiten y que los reciben. Por otro lado, en la mayoría de los sistemas de historia clínica se almacena localmente información clínica relativa al paciente, en muchas ocasiones a partir de los informes recibidos desde otros sistemas o dominios. De modo que mantener la total autonomía de los sistemas de historia clínica que se añaden a la federación puede significar, en algunas ocasiones, que una misma observación clínica se almacene en distintos sistemas.

Aún contando con esa posibilidad de duplicidad, las inconsistencias en valores son muy inferiores a las de otro tipo de información, como la demográfica, ya que aunque exista la posibilidad de duplicidad la información clínica habitualmente tiene muy pocos cambios y éstos se gestionan mediante mecanismos de versiones que facilitan la integridad de entidad. En la vista de información será necesario señalar el grado de duplicidad que se espera en cada tipo de observación para que los agentes tomen las medidas oportunas para resolver los conflictos de entidad.

Cada dominio debe tener agentes autónomos de gestión de observaciones que permitan, previa autorización, la gestión de las locales. En la arquitectura también estarán presentes agentes federados con la misma misión, que permitan la composición de observaciones a partir de otras más simples que pueden estar almacenadas en distintos

dominios, y que resolverán los conflictos de entidad, en caso de que éstos existan. Es evidente que estos agentes de observación federados además de interactuar con sus homónimos autónomos deberán interactuar, a través de vínculos, con otros tipos de agentes, incluso de distintos flujos de trabajo, para resolver problemas como la gestión de identidad y credenciales del actor en cada dominio, la semántica de la información, el idioma, los posibles conflictos de entidad, etc. La figura 22 presenta algunos de los principales agentes que están directamente relacionados con estas actividades.

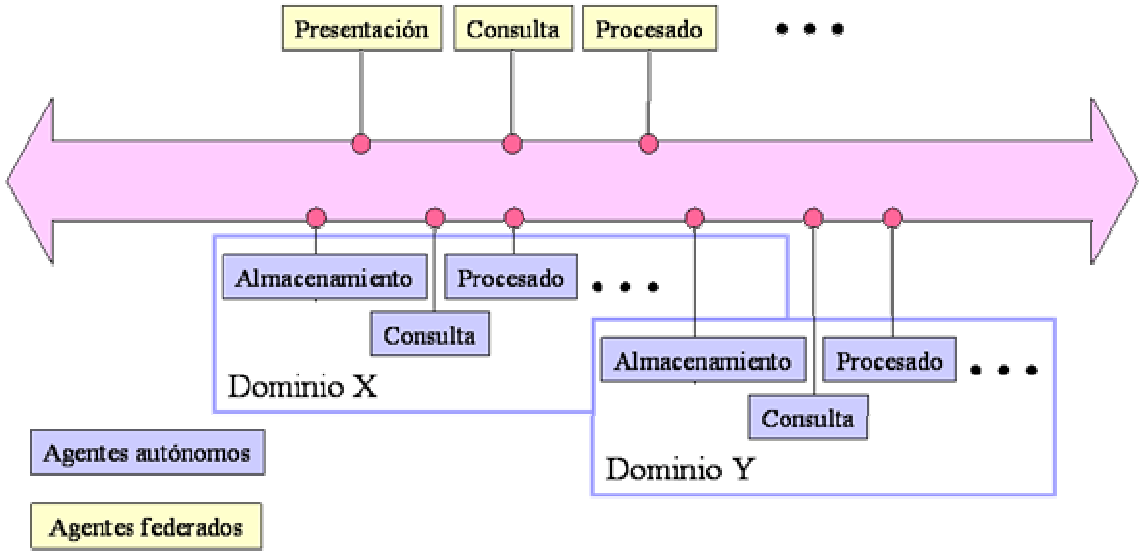


Figura 22. Principales componentes del flujo de trabajo de gestión de información clínica.

La norma señala que la mayoría de los datos de salud disponibles en una organización sanitaria son resultados de actividades clínicas llevadas a cabo por actores del sistema. Los procedimientos de adquisición de las observaciones, que son por supuesto muy variados, se pueden considerar también dentro de este grupo de actividades y será necesario incluir agentes que soporten estas tareas. Se podrán crear relaciones entre los datos clínicos y las actividades de usuario lo que puede contribuir a lograr algunos objetivos adicionales como son:

- La mejora de la calidad asistencial y efectividad del tratamiento al paciente, ya que los actores sanitarios tienen más detalles sobre el contexto en el que determinada observación clínica fue tomada.

- La posibilidad de la monitorización de costes y gestión de calidad dentro de la organización, relacionando los resultados obtenidos con las actuaciones realizadas.

Por otro lado, la arquitectura no puede estar pensada únicamente para la recuperación y presentación de observaciones, sino que tiene que ser el soporte para tareas más complejas; en este sentido se puede hablar de actividades de gestión del conocimiento clínico, que podríamos descomponer en tareas de recuperación, creación y almacenamiento. Las primeras se refieren a la recolección personalizada y contextualizada de las observaciones clínicas. Las actividades de creación de conocimiento clínico se refieren al uso de cualquier herramienta de ayuda a la decisión que se le ofrezca al usuario. Y las últimas a la correcta anotación y almacenamiento del conocimiento generado tras la intervención del usuario para su uso posterior. Es necesario señalar que la naturaleza y origen de las observaciones clínicas es muy variada, desde el análisis de laboratorio realizado a un paciente, hasta la medicación que actualmente está tomando, y de hecho es muy frecuente que éstas sean de carácter textual, lo que implica una dificultad añadida a la hora del procesado. Una correcta anotación semántica de las observaciones clínicas facilitará el desarrollo de componentes y aplicaciones de gestión del conocimiento clínico. En este sentido las actividades del flujo de información clínica están muy relacionadas con las de gestión semántica.

Por último, sólo observar la fuerte relación de estas actividades con las de comunicación. Por un lado para soportar algunos procedimientos de adquisición de observaciones clínicas, y por otro cuando en la gestión de conocimiento clínico se utilice información que provenga de sistemas externos y que incluso puede tener naturaleza no clínica.

6.3.4. Flujo de gestión de actividad

Dentro de la arquitectura hay que considerar y soportar todas las actividades requeridas por los usuarios para cubrir el conjunto de prestaciones de la organización. Algunas de estas actividades estarán relacionadas con tareas asistenciales y otras serán de propósito científico, organizativo o administrativo. Todos los aspectos relativos a la gestión de actividades son cruciales en la organización sanitaria. La integración, monitorización y optimización del flujo de trabajo para coordinar a todos los actores y componentes

implicados en una actividad es un objetivo principal para mejorar la calidad asistencial y de las actividades de organización y gestión. Las tareas que pertenecen a este flujo de trabajo implementan las funcionalidades necesarias para la coordinación de estas actividades.

La ejecución de una actividad concreta puede requerir unas precondiciones en el sistema, es decir el cumplimiento de ciertas condiciones de contorno, y/o la ejecución de otras actividades previas. A su vez una actividad provocará un efecto, al cambiar las condiciones del sistema y podrá disparar la ejecución de otras tareas. La formalización de actividades complejas como secuencias de actividades más elementales, mutuamente dependientes y relacionadas, proporcionan la definición de protocolos para llevar a cabo propósitos clínicos o de organización. En el caso de propósitos clínicos es frecuente la utilización por parte del personal sanitario de guías clínicas y planificaciones relativas a casos de salud o pacientes concretos (planes estándar o planes de paciente). Casi cualquier tipo de actividad puede ser descrita a través de un modelo que refleje las interacciones de distintos actores y componentes de la arquitectura durante las fases o estados del ciclo de vida de la actividad.

En el ciclo de vida de una actividad suelen estar implicadas las siguientes entidades:

- Un actor que solicita una o varias actividades.
- Uno o varios proveedores que ejecutan las actividades requeridas por el actor solicitante.
- Un gestor de las actividades o proveedor de servicios.

Así mismo el ciclo de vida de una actividad puede describirse del siguiente modo.

- Debido a necesidades locales o por una petición recibida desde fuera, un actor puede decidir que es necesario que otro u otros ejecuten determinadas actividades.
- Para cada conjunto de actividades cuya ejecución sea gestionada por determinado gestor debe enviarse una petición formal al mismo. Así pues el actor primero debe identificar qué gestor es el encargado de estas actividades y qué información necesita para llevarlas a cabo. Al realizar la solicitud de la ejecución de la actividad entre otras cosas habrá que indicarle qué actividades

específicas se solicitan, las razones y toda la información adicional que sea necesaria.

- Cuando el proveedor de servicios recibe la petición, informa al actor que solicita la actividad de la aceptación de la petición y puede también enviarle o solicitarle información adicional.
- Si alguna de las actividades solicitadas al proveedor de servicios no puede ser llevada a cabo, éste puede decidir reenviar la petición a otro e informará al actor que solicitó la actividad de este hecho.
- Una vez que el proveedor de servicios ha planificado una actividad y tiene una aproximación de cuándo será ejecutada se lo puede comunicar al actor que la solicitó.
- Incluso después de la planificación de una actividad se pueden realizar modificaciones por parte del actor que la solicitó o del gestor. Estas modificaciones deben ser negociadas entre ambos en función de sus requisitos particulares.
- En cierto momento la actividad será llevada a cabo por el proveedor. Durante la ejecución de la actividad puede intercambiarse información entre el actor solicitante y el proveedor.
- El proveedor puede requerir actividades adicionales de las que deberá informar al solicitante.
- Cuando todas las actividades relacionadas con la solicitud sean llevadas a cabo y se elabore un informe con los resultados de la actividad, la petición inicial se considera completada y se envía una comunicación formal al actor solicitante que reconocerá la recepción de la misma.

Haciendo hincapié de nuevo en el modelo federado, hay que destacar que los actores implicados en una actividad pueden pertenecer a distintos dominios y las actividades elementales que conforman actividades más complejas pueden ejecutarse en cualquier dominio. Tal y como está enunciado este flujo de trabajo no hay ningún problema en hacer esta consideración.

Esta forma en que la norma define y maneja el concepto de actividad es muy próxima al dominio de usuario, es decir que se refiere a actividades identificadas y solicitadas por

los actores del sistema. Pero es necesario añadir la consideración de que durante el ciclo de vida de una actividad además de los usuarios varios componentes de la arquitectura tendrán que interactuar. Esta interacción se llevará a cabo a través de los vínculos que se establezcan en sus interfaces. Por otro lado, estos vínculos pueden ser dinámicos, en el sentido de que la colaboración entre un conjunto de componentes determinado puede necesitarse únicamente en ciertas fases de la actividad, estableciéndose y liberándose los vínculos entre componentes cuando sea necesario. Los vínculos pueden ser complejos, de este modo se pueden establecer relaciones entre más de dos componentes siendo el objeto vinculador el que gobierna las interacciones implicadas. Esta gestión de interacciones se realiza según los requisitos y objetivos de la fase de la actividad en la que se crea el objeto vinculador y a los comportamientos y contratos de entorno establecidos en las interfaces que vincula. Las tareas que forman parte de este flujo de trabajo pueden incluir facilidades para la creación y liberación de vínculos en distintas fases de la actividad y la gestión de objetos vinculadores. La gestión de la ejecución de las distintas fases de las actividades será responsabilidad de los componentes de la arquitectura implicados.

Se necesitan componentes en la arquitectura que faciliten la gestión del ciclo de vida y de los distintos estados o etapas de una actividad. Estos componentes tendrán al menos las tareas de:

- Proveer mecanismos para permitir las interacciones entre componentes durante las distintas fases del ciclo de vida de una actividad, desde la petición inicial hasta el informe final, dando soporte a aspectos funcionales y de estructura de la información; entre otros, la instanciación de objetos vinculadores entre componentes.
- Proveer un repositorio común, accesible a los componentes de la arquitectura que lo necesiten, que contenga información sobre cada actividad posible dentro de la organización. Servirá de soporte para los requisitos clínicos y de organización de los diversos actores y para las actividades de cálculo de costes y calidad del sistema.
- Proveer de elementos que faciliten la creación de protocolos, guías clínicas y planes personalizados utilizando un mecanismo eficiente de especificación formal.

Algunas de las funciones básicas identificadas como pertenecientes a este flujo se muestran en la figura 23.

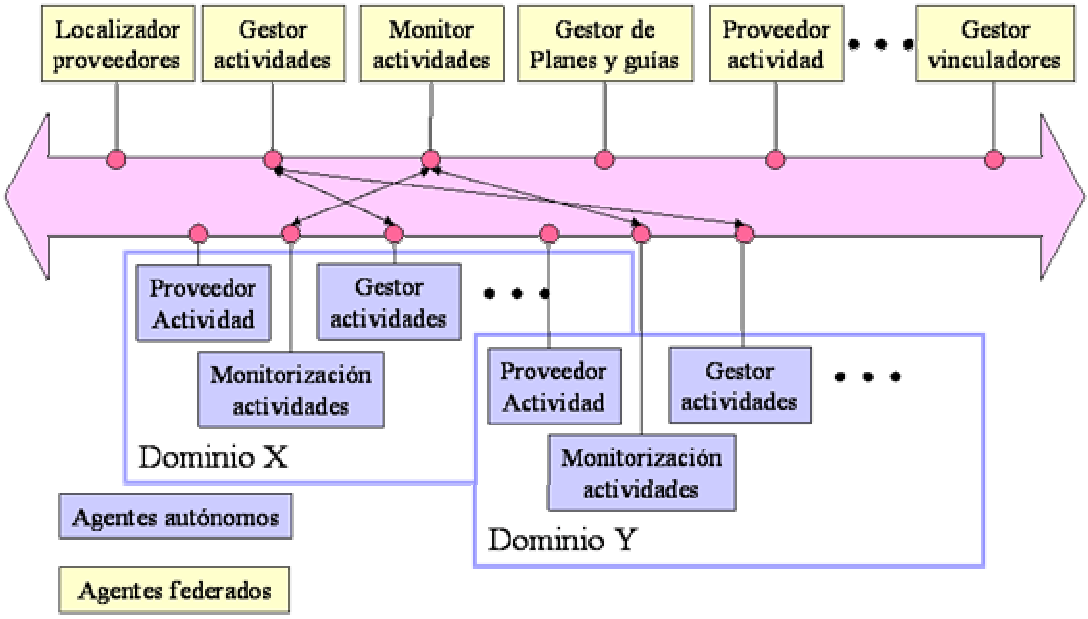


Figura 23. Principales componentes del flujo de gestión de actividad.

6.3.5. Gestión de recursos de la organización sanitaria

Cuando en la vista de empresa HISA se refiere a recursos, éstos representan los elementos necesarios dentro de la organización para el desarrollo de las actividades de usuario, sean éstas clínicas o de otra naturaleza. Se consideran recursos los materiales y equipos, el personal e incluso localizaciones físicas. Es evidente que dada la importancia de estos recursos, que no son propios de los sistemas de información sino de la organización, en las distintas actividades que pueden acontecer en el escenario sanitario será necesario proporcionar mecanismos para permitir la definición y recuperación de información sobre los recursos disponibles, así como para la gestión de los criterios y reglas para la utilización de éstos durante determinadas actividades.

En cuanto a la integración conforme al modelo federado, sólo remarcar que la gestión de recursos generalmente seguirá siendo una tarea que cada sistema autónomo querrá seguir realizando localmente. Las decisiones de los componentes autónomos en cuanto al uso de los recursos de un sistema concreto serán determinantes para la planificación de actividades dentro de la arquitectura.

6.3.6. Actividades de seguridad y gestión de recursos

En los sistemas sanitarios, aparte de los mecanismos de seguridad de cualquier sistema de información, hay que realizar consideraciones adicionales debidas a la especial sensibilidad de la información que se maneja, incluyendo factores éticos y legales.

Se deben considerar dos aspectos fundamentales y complementarios en la arquitectura sanitaria: la seguridad de los datos gestionados, y que el uso de la información almacenada y el acceso a las prestaciones ofrecidas sean realizados sólo por usuarios o agentes software autorizados [141].

En cuanto al primer aspecto, se refiere a los criterios y mecanismos según los cuales los datos deben ser gestionados en la arquitectura para asegurar los niveles de disponibilidad y protección adecuados. Entre otros deben incluirse procedimientos de encriptación de la información así como el uso de mecanismos y dispositivos para la correcta autentificación de individuos, aspectos que pueden estar muy próximos a las características técnicas concretas de la arquitectura y en muchos casos la plataforma de integración elegida facilitará estas prestaciones de forma transparente. Como consecuencia la norma HISA considera estas características de seguridad fuera del estándar.

En cuanto a lo segundo, un requisito indispensable en la arquitectura es soportar mecanismos para definir criterios y reglas de acuerdo a las cuales los usuarios serán autorizados a utilizar recursos del sistema para realizar determinadas actividades según las credenciales de los mismos.

Como muestra la figura 24 y sin profundizar en la complejidad de las actividades de seguridad éstas podrían clasificar del siguiente modo:

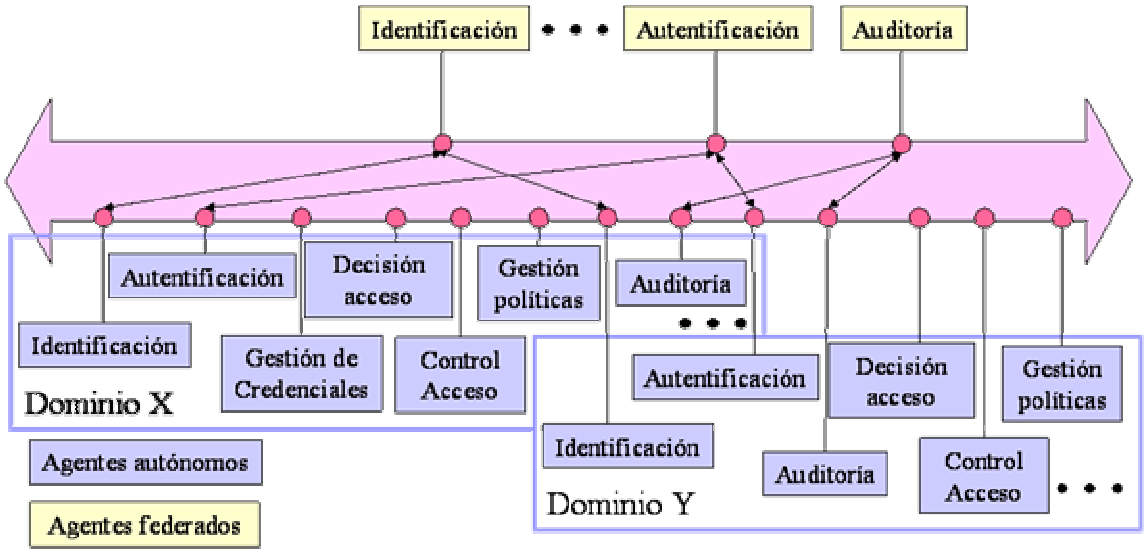


Figura 24. Principales componentes de seguridad en la arquitectura sanitaria

- **De identificación y autentificación**: para comprobar que un usuario es quien dice ser, bien pidiendo identificación directa al usuario (por ejemplo en forma de nombre de usuario y contraseña) o bien con mecanismos que hagan transparente esta tarea al usuario, por ejemplo mediante el uso de certificados digitales. Este es un problema que en muchas ocasiones está ya resuelto por la tecnología de integración utilizada.
- **De obtención y gestión de credenciales**: las credenciales son conjuntos de atributos que especifican las características del usuario en un sistema, habitualmente en forma de privilegios, capacidades y roles. Normalmente no son obtenidas directamente del usuario, sino que se obtienen una vez identificado y autenticado éste y de forma totalmente transparente al mismo. La arquitectura debe facilitar la gestión de estas credenciales y la obtención de las mismas cuando sea necesario.
- **De autorización de acceso a recursos y gestión de políticas**: proporcionan mecanismos para determinar si un agente dado puede realizar o no determinado acceso a un recurso. Se basa en las credenciales presentadas y en las políticas de acceso previamente especificadas para cada recurso (o conjunto de recursos). También es necesario proveer de mecanismos que faciliten a los usuarios la gestión de las políticas de acceso.

- **De control de acceso**: aplican las decisiones tomadas en los mecanismos anteriores asegurándose de que sólo los agentes autorizados acceden a cada recurso.
- **De auditoría**: registran información sobre la utilización que se ha hecho de cada recurso de la arquitectura: instante de tiempo, tipo de acceso, usuario, etc, y generan informes con los eventos ocurridos. Sirven de soporte para los mecanismos de protección de datos y para cumplir las legislaciones vigentes en cada dominio sobre estos aspectos.

Algunas de estas actividades implican directamente al usuario mientras que otras son transparentes al mismo y en ocasiones fuertemente transversales, es decir que son funcionalidades que se invocan a lo largo de todo el ciclo de vida de alguna actividad de usuario.

Como señala el estándar, las plataformas tecnológicas actuales pueden solucionar algunos de estos problemas, que podrían quedar fuera de la definición específica de una arquitectura para sistemas sanitarios. En cualquier caso son actividades que es necesario considerar en un sistema de estas características.

Dada la complejidad de todas estas actividades, aunque HISA trata juntas las tareas de gestión de usuarios y autorizaciones en esta Tesis se considera que es preferible separarlas para dar mayor flexibilidad a la arquitectura: por un lado las tareas de gestión de usuarios y credenciales y por otro lado las de gestión de recursos y autorizaciones, aunque estén muy relacionadas y en muchos sistemas se implementen juntas. A su vez se ha distinguido entre recursos de la arquitectura y recursos de la organización, en lo que se profundizará en los próximos puntos.

Actividades de gestión de usuarios y credenciales

Las actividades de gestión de usuario y credenciales se centran básicamente en registrar, identificar y validar a los usuarios que pueden acceder al sistema y gestionar el papel o rol que juegan dentro del mismo. Estas tareas son especialmente complejas en escenarios federados. Al igual que se señaló con la identificación de pacientes, la identificación de usuarios debe seguir siendo una tarea local. Cada sistema tiene registrados sus propios usuarios y por tanto, si se quiere evitar el modelo centralizado, un usuario registrado en el sistema debe estar registrado en cada uno de los que lo componen, o al menos en todos aquellos a los que tiene que acceder. Esto da lugar a una

alta duplicidad de la información de identificación de usuarios con la consiguiente necesidad de resolver los conflictos de entidad que puedan surgir.

Las credenciales proporcionarán al usuario capacidades para utilizar o acceder a determinados recursos, es decir identifican el rol de ese usuario en el sistema; por tanto, para un mismo usuario éstas son diferentes en cada uno de los sistemas de la federación. Es evidente que en un modelo totalmente federado cada sistema gestiona las credenciales de sus usuarios de forma autónoma.

Gestión de recursos de arquitectura y autorizaciones

Cada componente de la arquitectura se puede describir en términos de un conjunto de funcionalidades que sólo pueden ser utilizadas por otros componentes según reglas de autorización previamente establecidas. Del mismo modo, cada elemento de información en el sistema sólo está disponible para los usuarios con las credenciales apropiadas y para determinados usos. Aunque en un principio esto pueda parecer sencillo, la realidad demuestra que no es así: la gran cantidad de componentes de la arquitectura, de variedad de información manejada por los sistemas, la diversidad de usuarios y las interrelaciones usuario-sistema provocan que sea necesario desarrollar complejas políticas de acceso que introduzcan algo de orden. Estas políticas básicamente relacionan recursos, operaciones y usuarios según estructuras jerárquicas que representan las posibles relaciones entre ellos. La arquitectura debe proveer de componentes que faciliten la gestión de estas políticas de acceso en las que puede ser necesario considerar parámetros de naturaleza muy diversa.

En un modelo federado, esta gestión de políticas estará en manos del responsable del recurso y por tanto será una gestión local que por supuesto puede ser muy distinta de unos sistemas a otros. Pero los modelos que se pueden encontrar en una organización sanitaria son diversos y complejos. Si se piensa, por ejemplo, en una observación clínica determinada a la que se quiere acceder a través de componentes de gestión de historia clínica, centrados en el paciente, es comúnmente aceptado que es el paciente el que debe autorizar el uso de su información personal. Es necesario desarrollar mecanismos que en los sistemas de historia clínica faciliten la labor a los pacientes para personalizar las autorizaciones de acceso a sus datos personales, funciones que actualmente se encuentran implementadas en muy pocos sistemas. Sin embargo si la misma observación se consulta de forma despersonalizada, es decir sin conocer el paciente al que pertenece, a través de componentes con otros fines, por ejemplo como fuente para la

realización de estudios epidemiológicos, la decisión de acceso a estos recursos puede estar en manos de otros actores del sistema distintos al paciente. Este caso simplemente intenta reflejar el hecho de que para un mismo recurso pueden existir distintas políticas y, lo que es más complicado, distintos responsables de la gestión de las mismas según el uso y el contexto. Por otro lado, las implicaciones de la legislación referente a la gestión de información de carácter personal deben ser tenidas en cuenta al analizar los requisitos de la arquitectura para soportar estas actividades.

Gestión de auditoría

Las actividades de auditoría se encaminan al cumplimiento de la legislación vigente en materia de seguridad e integridad de datos de carácter sensible. Se consideran aquí tareas como la gestión de informes sobre el uso de los datos registrados o sobre los accesos a los sistemas, así como los mecanismos de gestión de versiones de la información almacenada. Aunque HISA no considera estas actividades por separado, sí indica que son necesarias y que el modelo de información debe incluir los elementos que las soporten. En esta Tesis se presentan como un conjunto de actividades independientes ya que tienen suficiente entidad y que, aparte de las características comunes a todas las clases HISA, se necesitará un modelo de información propio y la especificación detallada de estas tareas en la vista computacional.

Consideraciones de federación

La gestión de todas estas tareas es especialmente compleja en escenarios federados, donde cada uno de los sistemas puede considerarse un dominio de seguridad autónomo de manera que gestiona sus propios usuarios, credenciales, políticas de acceso,etc.

Al interconectar varios dominios de seguridad autónomos deberían cumplirse una serie de requisitos básicos:

1) Cada dominio define y aplica sus propias políticas de seguridad de forma totalmente autónoma; el hecho de pertenecer a una federación no alterará su funcionamiento local.

2) Incluir un dominio de seguridad en la federación debe ser sencillo y no debe suponer cambios en los sistemas que ya formaban parte de la federación.

3) El acceso a cualquier recurso de un dominio debe estar protegido por sus propios mecanismos de decisión de acceso, aunque se acceda desde otro dominio.

4) Los mecanismos de seguridad deben ser transparentes al usuario, salvo quizá para identificarse al iniciar una sesión de trabajo dentro del sistema. En cualquier caso, para realizar la identificación y autentificación de usuario en los distintos dominios de la arquitectura no se deben solicitar más datos que los requeridos en esta primera identificación.

5) La obtención de credenciales para acceder a cualquier dominio debe ser transparente al usuario.

La arquitectura de integración que se diseñe dependerá fuertemente del modelo de confianza elegido. Así puede realizarse la integración en el nivel de identificación y autentificación. En este caso se usan los procedimientos de identificación y autentificación y de obtención de credenciales locales a cada sistema, no confiando en las credenciales que vengan indicadas desde un dominio externo. Para cumplir los puntos 4 y 5 y conseguir que estos procedimientos, llevados a cabo a través de varios dominios, sean transparentes al usuario se podría usar un agente proveedor de identidad.

La misión de este componente sería obtener los datos necesarios, en el formato adecuado, para proceder a la identificación del usuario en cada uno de los dominios donde se encuentren los recursos a los que éste pretenda acceder. Este agente se apoyará en los componentes gestores de semántica para realizar las conversiones necesarias entre los modelos de identificación del usuario en cada sistema. Con este modelo de confianza los mecanismos de identificación y autentificación, de obtención de credenciales y de decisión de acceso se mantienen en cada sistema exactamente igual que antes de pertenecer a la federación asegurándose así que se cumplen el resto de condiciones impuestas al sistema.

Otro modelo de confianza sería aceptar como válidas las credenciales que se envíen desde determinado agente, al que llamaremos gestor de credenciales. En este caso la identificación y autentificación del usuario se haría sólo una vez y el gestor de credenciales obtendría las credenciales necesarias para acceder a recursos en cualquiera de los sistemas de la federación. Así la integración se hace en el nivel de decisión de acceso y no en el nivel de identificación. Cumplir el punto 3 antes señalado para los componentes de control de seguridad dentro de la arquitectura implica que deben seguir siendo los agentes de decisión de acceso de cada sistema los que se encarguen de tomar las decisiones de acceso dentro de su dominio de seguridad. Para cumplir los puntos 1 y

2 no se deben realizar cambios en estos agentes autónomos al integrarlos en el sistema, ni les debe afectar el que se añadan nuevos dominios de seguridad.

6.3.7. Gestión de semántica

Los requisitos especificados en la vista de empresa para las actividades relacionadas con la semántica del sistema indican que la arquitectura debe proveer de facilidades para los usuarios que cubran la gestión de:

- Los distintos sistemas de clasificación y tipos utilizados para los conceptos del sistema de información.
- Los vocabularios comunes que cubren el conjunto de términos que las aplicaciones necesitan y utilizan para describir el dominio de la aplicación.
- Las dependencias y reglas que pueden existir entre conceptos relacionados y entre instancias o individuos en función de los valores concretos de sus atributos.
- Las reglas que indican cómo instanciar los conceptos del dominio.
- Establecer las correspondencias de elementos del modelo de información de la arquitectura a los mensajes y vistas requeridos para soportar interacciones con otros sistemas o entre componentes en distintos dominios de la organización. Normalmente estos mensajes y vistas se basan en estándares de comunicación y modelos de datos normalizados.

Como consecuencia se identifican dos conjuntos de requisitos que deben ser soportados por la plataforma de integración:

- Para cada actividad de usuario la gestión de diccionarios, clasificaciones y reglas adoptadas por la organización para los procesos implicados en la misma.
- De forma global la gestión de dependencias y relaciones entre conceptos utilizados con propósitos clínicos, organizativos, de gestión, o cualquier otro, y su conversión a formatos conformes a otros estándares de interés.

Como se viene señalando, el modelo que el estándar plantea es definir conjuntos modulares, consistentes y autosuficientes de componentes, cada uno de los cuales proporciona las prestaciones necesarias para alguna de las actividades de usuario que se han identificado a lo largo de la vista de empresa. Como cada una de las actividades de

usuario necesitará cierta gestión semántica, el conjunto de agentes que soporta determinada actividad debe incluir estos componentes semánticos para evitar dependencias de agentes externos y aumentar la modularidad de la arquitectura. De este modo las actividades presentadas en este grupo son adicionales a las propias de cada conjunto y permiten mejorar las prestaciones de gestión semántica; algunas de las más importantes se presentan en la figura 25. Así, la fuerte transversalidad de las actividades de gestión semántica provoca que el desarrollo de componentes para estas actividades sea reutilizado en las demás, de manera que un desacoplo total de las mismas es prácticamente imposible.

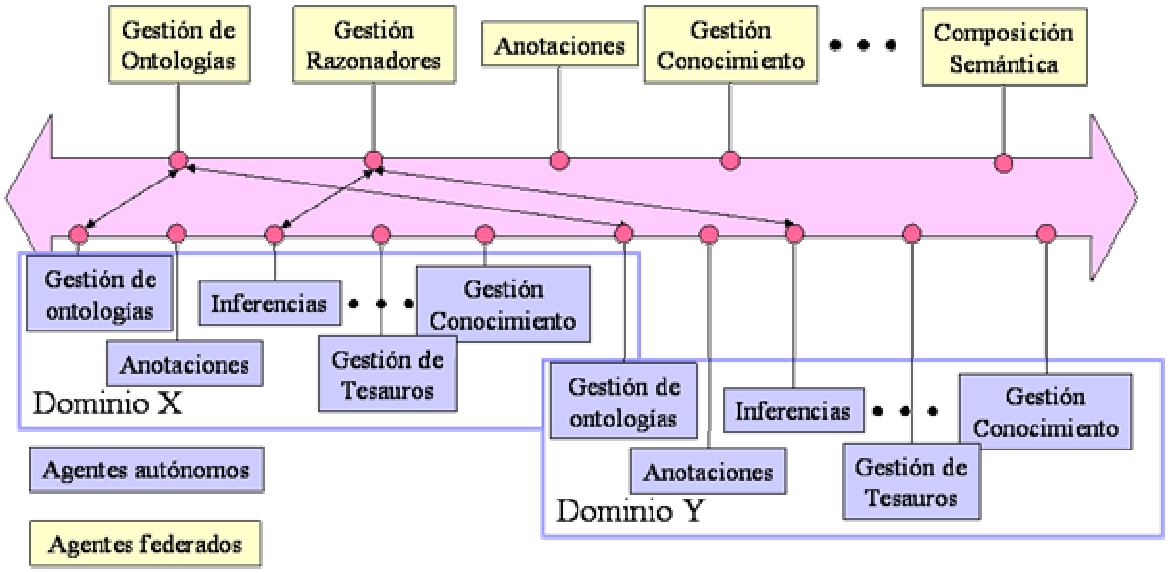


Figura 25. Principales componentes de gestión de semántica

En esta Tesis se amplían los requisitos de actividades semánticas recogidos en la vista de empresa al considerar que los avances realizados en cuanto a gestión semántica deben ser tenidos en cuenta al desarrollar un estándar de esta envergadura. Ante todo a partir de ahora además de referirnos a sistemas de clasificación, códigos o diccionarios, como la norma indicaba, hablaremos también de la **gestión de ontologías**.

La ventaja es que el concepto de ontología, ampliamente aceptado por la comunidad científica, tiene una mayor capacidad de expresar semántica. Ya se introdujo el término ontología en capítulos anteriores, pero recordemos que es el conjunto de conceptos, y relaciones entre éstos, dentro de un dominio. Además puede incluir también el conjunto de individuos del dominio.

Como se indicó anteriormente la gestión semántica es especialmente necesaria en una arquitectura que pretenda seguir el modelo federado, ya que uno de los mayores problemas a resolver es la reconciliación semántica de los sistemas que se van a integrar, que debe hacerse sin modificar cada uno de ellos. Es pues imprescindible que en la arquitectura existan componentes que ayuden a solucionar estos conflictos. Por otro lado los componentes basados en ontologías y gestión del conocimiento pueden ser piezas clave para el desarrollo de aplicaciones más complejas dentro de la arquitectura sanitaria.

La figura 25 presenta algunas de las principales funciones de gestión de semántica útiles en la arquitectura, que podrían resumirse como:

- **Gestión de tesauros**: se incluyen en este grupo todos los componentes tradicionales de gestión de sistemas de términos y sistemas de codificación.
- **Gestión de ontologías**: estos componentes gestionan repositorios de ontologías, permitiendo a los usuarios añadir y eliminar conceptos y relaciones, facilitando la consulta, la fusión de ontologías, etc. También pueden proveer de facilidades para hacer la correspondencia o fusión de distintas ontologías, por ejemplo de la ontología local de un componente a la común de la federación. Las tareas de correspondencia y fusión pueden ser manuales, automáticas o semiautomáticas. Es importante destacar que se podrían considerar dos tipos de usuarios de la arquitectura; los expertos en tecnologías de la información y el conocimiento y los expertos del dominio. Lo más deseable, y complicado, es desarrollar aplicaciones de gestión de ontologías que puedan ser manejadas por los segundos y se necesite la mínima intervención de los primeros. De este modo la definición de nuevos conceptos y relaciones debe estar en manos de los expertos del dominio. En el terreno de la historia clínica electrónica esta idea ha sido ya utilizada en la norma EN13606-2, donde se utiliza el concepto de arquetipo que, como se presentó en el capítulo dedicado a métodos y modelos, pretende facilitar al experto del dominio la definición de ontologías. En particular los componentes de gestión de arquetipos para historia clínica pertenecerían a este tipo de elementos dentro la arquitectura, pero al considerar una arquitectura de sistemas sanitarios se necesitará manejar ontologías relativas a otros dominios semánticos. Cuanto más abierto sea el modelo de componentes de gestión de ontologías más fácil será la integración de distintos dominios semánticos.

- **Gestión de razonadores**: serían agentes del sistema que ayudan a definir reglas de inferencia y a ejecutarlas cuando sea necesario. En muchas ocasiones son el fundamento para el comportamiento proactivo de los componentes de la arquitectura, ya que permiten que al observar algún cambio en las condiciones del entorno los componentes puedan, de forma autónoma, ejecutar alguna acción concreta. Al igual que ocurre con la gestión de ontologías, es deseable acercar al experto del dominio la gestión de estas reglas de inferencia.

- **Gestión de anotaciones**: pueden existir agentes en el sistema que ayuden a anotar la información almacenada en el sistema de información para añadirle nuevo conocimiento. Cuando hablamos de la gestión de información clínica ya señalamos que era importante poder almacenar el conocimiento generado tras la actividad asistencial con la suficiente integridad semántica para poder recuperarlo más tarde. Del mismo modo ocurrirá con la información no clínica. Es también importante destacar en este punto el hecho de que la mayor parte de la información clínica es de tipo texto libre, de manera que estos componentes suelen estar ligados a otros encargados del procesamiento del lenguaje natural, sobre todo cuando se pretende que las anotaciones se realicen de modo automático o semiautomático.

- **Gestión de conocimiento**: ya se habló de la gestión de conocimiento clínico al especificar el flujo de información clínica, pero es evidente que en una arquitectura que intente cubrir todas las necesidades de una organización sanitaria es necesario tener en cuenta información de otro tipo y relaciones entre piezas de información muy distintas. Dentro de estas actividades pueden existir aplicaciones para la gestión eficiente de recursos, análisis de rendimiento y costes de la organización, planificación, etc.

- **Composición semántica de componentes**: el número de componentes disponibles y las posibilidades de nuevas aplicaciones crecerá a medida que la arquitectura se vaya desplegando e implantando. Se necesitarán facilidades para crear nuevos componentes y aplicaciones a partir de los disponibles y, por supuesto, para publicar y poner a disposición de los demás las nuevas prestaciones desarrolladas. Cuanto más crezca la carga semántica en la especificación de interfaces más eficientes y automáticas serán las tareas de descubrimiento, invocación y composición. Como han mostrado algunas de las

líneas de investigación actuales, presentadas en el capítulo de técnicas de semántica, la panacea es conseguir una descripción semántica de componentes suficientemente rica para que a partir de la definición de un objetivo la invocación y composición de los mismos sea automática. Si esto se lograra no se requeriría la normalización de componentes, sino únicamente la normalización de la forma de describir su funcionalidad, su interacción con los demás y la información manejada.

6.3.8. Gestión de comunicaciones

Este grupo de actividades es necesario porque en una organización sanitaria los modelos de actividad y de interacción entre usuarios cada vez son más diversos. Aunque la arquitectura se diseñe para unos mecanismos de interacción determinados, debe estar abierta al desarrollo de otros nuevos. Por otro lado la integración entre sistemas sanitarios puede hacerse de forma totalmente acoplada, incluyendo en la arquitectura todos los componentes que soporten las actividades de dichos sistemas, aunque también será frecuente encontrar integraciones más ligeras. En estos casos se desarrollan componentes específicos que sirven de mediadores entre la arquitectura y el sistema externo.

Los mecanismos para gestionar estas comunicaciones se podrán identificar como actividades de usuario específicas, pero además existirán componentes que, de forma transparente al usuario, realicen tareas relacionadas con estas actividades. Esto muestra el carácter transversal de las operaciones de comunicación y la alta reutilización de estos componentes en la arquitectura sanitaria.

HISA no indica en ningún momento las dimensiones y la cobertura que una arquitectura de esta naturaleza debe tener. De manera que un sistema desarrollado para un hospital puede ser conforme a HISA, pero también la arquitectura sanitaria de una administración regional o nacional puede cumplir los requisitos necesarios. Es precisamente esta capacidad de escalabilidad una de las mayores ventajas del modelo. Los posibles escenarios para el desarrollo de las actividades, los usuarios, los mecanismos de comunicación y la tipología de aplicaciones aumentan a medida que se aumenta la cobertura de la arquitectura. Será necesario incorporar nuevos componentes que soporten estos nuevos requisitos y, por supuesto, reutilizar correctamente los ya existentes.

Considerando exclusivamente, y a modo de ejemplo, los modelos de interacción con carácter asistencial que podemos encontrar entre pacientes y profesionales de una organización sanitaria tendríamos, entre otros:

- Atención hospitalaria
- Atención ambulatoria
- Atención en urgencias
- Teleconsulta [142]
- Telemonitorización

Las aplicaciones que soportan las actividades de usuario tienen unos requisitos de comunicación, de calidad de servicio y de acceso a la arquitectura que dependen del modelo de interacción. En una sesión de teleconsulta, por ejemplo, se establecen flujos de datos en tiempo real entre dos terminales, uno de ellos suele estar en un domicilio y accederá a la arquitectura a través de una red de acceso residencial, como la RTB, mientras que el otro puede estar en un centro de salud y acceder a través de la red corporativa. Durante la telemonitorización de un paciente se puede requerir que sensores que porta el paciente envíen información y que existan componentes de la arquitectura que se encarguen de la gestión de alarmas. Si el diseño de la arquitectura subyacente es suficientemente modular y estructurado, muchos componentes podrán reutilizarse en diversas aplicaciones y se facilitará la integración de sistemas que fueron creados con distintos fines y que se basan en distintos modelos de interacción.

Como se mostró anteriormente, el flujo de gestión de actividad incluye los componentes que proveen de mecanismos para las interacciones funcionales y de información entre componentes durante el ciclo de vida de una actividad. Esta definición es muy amplia y podría considerarse que las actividades que se incluyen dentro de este conjunto que hemos añadido al modelo están ya cubiertas en este flujo. Sin embargo, aunque así fuera, la complejidad de estas tareas es tan grande que si se externalizan los mecanismos para gestionar las comunicaciones se aumenta en gran medida la escalabilidad del sistema.

6.3.9. Consideraciones de los sistemas de historia clínica única en la vista de empresa

Dentro de la arquitectura HISA es de especial interés identificar las actividades de usuario relacionadas con la gestión de la historia clínica de un paciente y analizar los vínculos entre estas actividades y las mostradas en la vista de empresa. Se puede comprobar que en realidad las actividades de gestión de historia clínica son de un nivel alto, es decir, son composición de actividades más básicas que se pueden corresponder a alguna de las ya mostradas en la vista de empresa de HISA. La figura 26 trata de reflejar esta idea de relación entre las actividades ya presentadas y las propias de la gestión de HCE.

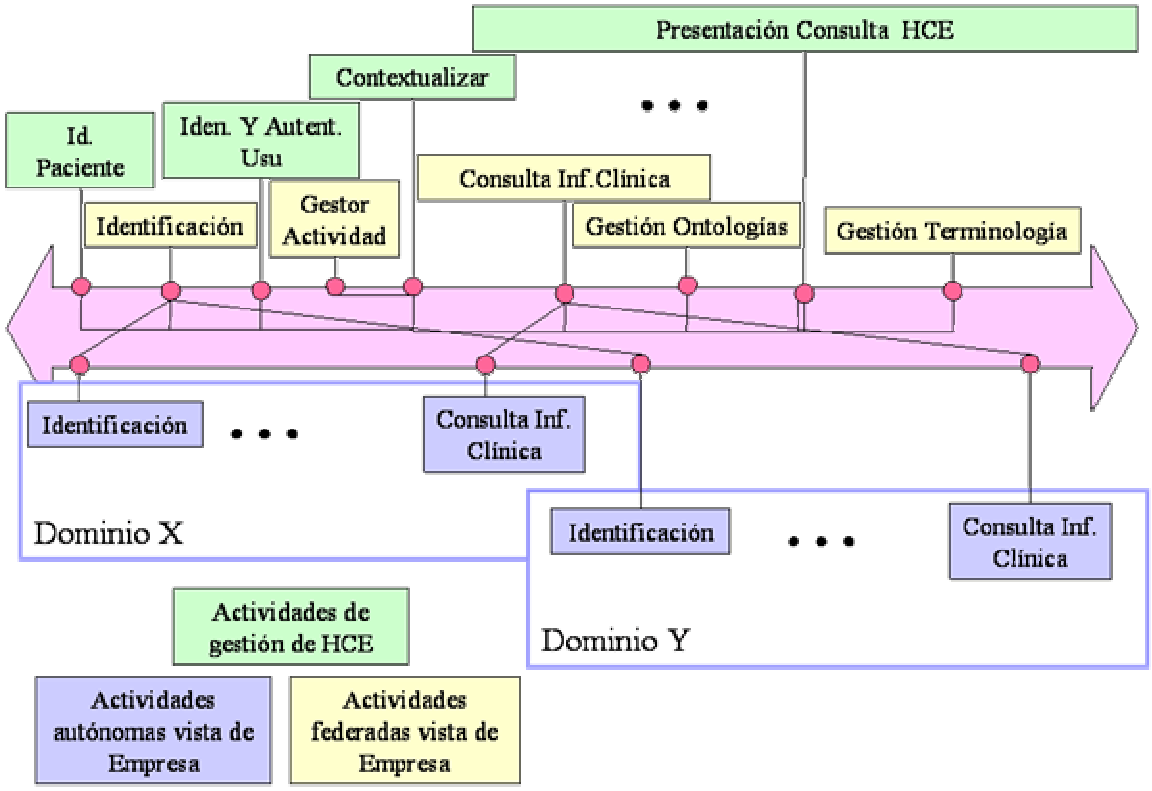


Figura 26. Representación de algunas actividades de gestión de HCE y su relación con actividades presentadas en la vista de empresa.

Volviendo a la definición de historia clínica, recordemos que es el conjunto de documentos relacionados con la asistencia sanitaria realizada a un paciente. Por tanto está orientada al paciente, que es el punto de conexión de toda la información que se va a gestionar. Así que hay que comenzar por identificar al paciente, tarea que, como vimos, formaba parte de las actividades del flujo del paciente.

Una vez identificado al paciente en todo el sistema es necesario recuperar la información que sobre el mismo hay almacenada. En una historia clínica tradicionalmente se puede distinguir entre información demográfica y clínica, aunque en conceptos más amplios de historia clínica podría considerarse información de otra naturaleza. En cuanto a la gestión de información demográfica, también puede considerarse como parte de las actividades del flujo del paciente. Estas actividades se utilizarán para la validación de los métodos propuestos en esta Tesis Doctoral.

Por otro lado la recuperación contextualizada de la información clínica de un paciente, la generación de nuevo conocimiento con las herramientas proporcionadas por el sistema más la intervención del profesional sanitario y por último el almacenamiento del nuevo conocimiento generado, pueden considerarse actividades relacionadas con la gestión de historia clínica y, como vimos, forman parte de las actividades del flujo de información clínica.

La gestión de la historia clínica por parte de los profesionales implicados en la atención al paciente se realizará dentro de una determinada actividad asistencial, y estará por tanto también relacionada con el flujo de gestión de actividad. De hecho una de las principales ventajas de considerarlo así es que identificar que cierta información de la historia clínica fue registrada durante una actividad en particular puede suponer un conocimiento añadido muy útil tanto para labores asistenciales como de gestión.

En cuanto al resto de actividades identificadas en la vista de empresa, también existen vínculos con la gestión de la historia clínica. El acceso a la información de un paciente debe estar regulado por mecanismos de seguridad y autorización, se debe asegurar la integridad semántica con las tareas de gestión semántica, los componentes de gestión de conocimiento son utilizados por las aplicaciones de soporte a la decisión clínica, la información puede ser enviada a otros sistemas y las aplicaciones de usuario que permiten la gestión de la historia clínica pueden utilizarse en distintos tipos de terminales y localizaciones gracias a los mecanismos de comunicación, etc.

De manera que diseñar los sistemas de historia clínica como componentes dentro de una arquitectura sanitaria más compleja permitiría que las aplicaciones desarrolladas exclusivamente para la gestión de la historia clínica pudieran hacer uso de componentes ya desarrollados como soporte de otras actividades identificadas en la vista de empresa. Así la gestión de la historia clínica puede incluir prestaciones más avanzadas y

centradas en un paciente concreto como por ejemplo acceso a fuentes de conocimiento externas, tratamiento personalizado y asistido, etc.

Del mismo modo, componentes que en su origen fueron desarrollados como soporte de las aplicaciones de gestión de historia clínica pueden utilizarse para el desarrollo de otras aplicaciones que provean de nuevas prestaciones, y que soporten actividades de distinta naturaleza, por supuesto siempre y cuando el diseño fuera conforme al modelo de arquitectura normalizado. Estas nuevas aplicaciones pueden tener cualquier carácter y pueden no estar orientadas a un paciente concreto: generación de conocimiento basado en la evidencia, herramientas de soporte a la decisión clínica, gestión hospitalaria, etc.

Para terminar señalar de nuevo el carácter de sistema federado al hablar de un sistema de gestión de historia clínica única. Si se incorporan distintos sistemas de historia clínica que se gestionan de forma autónoma es posible que haya información relativa al mismo paciente almacenada en varios de estos sistemas. Aplicaciones para la gestión de la historia clínica única deberían permitir al usuario manejar toda esta información de forma transparente al hecho de la distribución de la información en sistemas autónomos. Si cada sistema fue diseñado conforme a HISA y siguiendo el modelo ODP, el desarrollo de estas aplicaciones de historia clínica única se apoyaría en la reutilización de los componentes ofrecidos por cada sistema de historia clínica que se incluya en la federación y de los componentes de la arquitectura que proporcionen otras prestaciones necesarias como la integridad semántica, la identificación del paciente en toda la federación, la autentificación de usuario, etc.

6.3.10. Consideraciones finales y conclusiones

A lo largo de la definición de la vista de empresa se ha podido comprobar que aunque se intenten diferenciar actividades de usuario, éstas pueden estar relacionadas unas con otras de muy diverso modo. Esto indica que la reutilización de los componentes que soportan las aplicaciones de usuario será una gran ventaja, pero evidentemente sólo podrá llevarse a cabo si el diseño se hace correctamente.

En esta Tesis Doctoral se han realizado algunas aportaciones a la vista de empresa que incluyen.

- La reestructuración y ampliación de las actividades de usuario que se pueden identificar en una arquitectura de sistemas sanitarios.

- La definición detallada de las características de las actividades de usuario añadidas.
- Incluir allí donde es necesario matices que el modelo federado obliga a considerar.
- Reflejar el papel que las actividades de gestión de historia clínica juegan en la arquitectura sanitaria y remarcar que las aplicaciones que soportan estas actividades deben ser desarrolladas siguiendo el modelo ODP y las consideraciones de HISA para facilitar la incorporación en una arquitectura sanitaria.

6.4. Extensión de la vista de información

6.4.1. Introducción

HISA especifica que dentro de la arquitectura se debe integrar y organizar la información compartida por los agentes en un modelo de información consistente que soporte todas las necesidades de las actividades identificadas en la vista de empresa. La especificación debe estar basada en un modelo de objetos (aunque en este trabajo se habla de modelo de clases o modelo de referencia por considerarlo más apropiado) que sirva para estructurar adecuadamente la información que se identifique como importante en la especificación del conjunto de procesos de negocio, tareas y actividades indicadas en la vista de empresa.

Hay que recordar que el modelo de información es la base para lograr la integridad semántica, es decir los conceptos y relaciones que se presentan en la vista de información tienen que ser entendidos por los componentes de la arquitectura sin posibilidad de que exista confusión. El modelo de información es la ontología compartida del dominio. Precisamente uno de los principales problemas que se vienen señalando para conseguir la interoperatividad entre componentes siguiendo el paradigma de federación es que las aplicaciones que se desarrollan, y por tanto los agentes que las soportan, siguen distintos modelos de información ya normalizados o incluso utilizan modelos propietarios. Es importante destacar que esto no va a cambiar y que proponer como solución que todos los desarrolladores utilicen los mismos modelos de datos en sus sistemas no es realista. De este modo el modelo de información de la federación tiene que ser suficientemente flexible para permitir en todo caso el

establecimiento de equivalencias desde los modelos locales y la arquitectura debe proveer de componentes que faciliten estas tareas a los desarrolladores locales. Así no se obligará en ningún caso a cambiar los sistemas a integrar y se facilitará la integración en la federación.

Como se ha señalado anteriormente uno de los mayores problemas en el dominio sanitario es precisamente el dinamismo de los conceptos implicados en el mismo y el modelo dual facilita esta gestión. En el marco del punto de vista de información se puede considerar que este paradigma se traduce en que el modelo de información está dividido en clases pertenecientes a un modelo de referencia estable y clases dinámicas que básicamente restringen las clases del modelo de referencia para representar conceptos específicos. En los estándares de historia clínica estas clases dinámicas son las que se conocen como arquetipos. Las restricciones serán impuestas por los expertos en el dominio cada vez que se cree o modifique un concepto. En un modelo ontológico se puede considerar que tanto las clases del modelo de referencia como los arquetipos son conceptos dentro de la ontología del dominio, simplemente son de distinto nivel semántico: los primeros más próximos al dominio tecnológico y los segundos más cercanos al clínico.

Crear un nuevo modelo de referencia para la información manejada en HISA no parece aportar ningún beneficio de integración, ya que existen numerosos modelos de datos ya normalizados que podrían utilizarse. Simplemente habría que escoger entre los ya existentes alguno, o algunos en caso de que ninguno cubra por completo el dominio de HISA, y usarlos en la vista de información de la arquitectura.

Sin embargo existen varios factores que pueden influir en el interés de formalizar la vista de información dentro de HISA:

- Que el modelo de información debe servir de soporte a todas las necesidades de las vistas de empresa y computacional y que éstas pueden no estar cubiertas en los modelos ya normalizados.
- En relación con el punto anterior, al ser algunas de las actividades muy transversales es necesario que determinada información esté en muchas de las clases del modelo de información.
- Que la nueva ontología compartida debe ser suficientemente flexible para acomodar sin problemas cualquier modelo de información ya existente o que

pueda existir en el futuro. En un modelo federado la ontología tiene el papel de integración semántica y debe ser creada con esa finalidad.

- Que la propia ontología debe incluir mecanismos que faciliten tanto su extensión como las posibles operaciones de correspondencia con las ontologías manejadas localmente por los componentes de la arquitectura.

Si el modelo de información se considera la ontología de integración, y precisamente el objetivo es mantener la autonomía de los sistemas, hay que decidir cómo establecer equivalencias desde los distintos sistemas que participan en la federación, ya que evidentemente éstos no siempre han sido desarrollados conforme a esta ontología. Se pueden considerar dos mecanismos:

- **Ontología federada única:** la ontología federada crece a medida que se incluyen sistemas añadiendo el modelo de información de cada uno y estableciendo las relaciones entre conceptos. Así que HISA provee una ontología básica y en una organización concreta ésta crecerá incorporando todos los conceptos manejados en cualquiera de los sistemas y las relaciones entre éstos y los previamente existentes. De este modo lo que se consigue, una vez que en la plataforma se incluyen los componentes de gestión semántica, es que todos los sistemas sean capaces de entender a todos los demás porque tienen "el diccionario" necesario para establecer las relaciones entre sus conceptos y los manejados por el resto. El mayor problema sería el crecimiento desmesurado de la ontología, sobre todo teniendo en cuenta que al ir añadiendo los modelos de cada uno de los sistemas pueden crearse redundancias innecesarias para representar el mismo concepto. Esta circunstancia puede dificultar las tareas de razonamiento para encontrar las relaciones entre conceptos, especialmente en arquitecturas grandes y cuando los sistemas a integrar hayan sido desarrollados sin seguir estándares.
- **Correspondencias desde ontologías locales a la federada:** la ontología federada, normalizada en HISA, sólo se aumenta cuando aparece un concepto completamente nuevo, no recogido previamente en la misma. Es decir se evitan redundancias en el modelo federado. Cada sistema está obligado a utilizar este modelo de información común, de manera que se incorporan los componentes necesarios para realizar transformaciones a la ontología federada. La ventaja de

publicar el modelo de información como una ontología es la facilidad que se da a los sistemas locales para comprender el modelo federado y realizar estas transformaciones. Además se pueden proporcionar dentro de la arquitectura componentes semánticos que les ayuden en este proceso.

El esfuerzo de desarrollo del modelo de información de HISA debe ser por tanto realizar una profunda revisión de los modelos ya normalizados, seleccionarlos y extenderlos en la medida de lo necesario para presentar una ontología base que irá evolucionando de forma distinta en cada arquitectura que aplique esta norma. Además es lógico pensar que dado que HISA es un estándar promovido por el CEN, ante igualdad de prestaciones se seleccionarán los modelos normalizados por este organismo para esta ontología base. Así, por ejemplo, para modelar la información manejada por los componentes encargados de la gestión de información clínica puede preferirse utilizar la ontología presentada en el EN13606-1, antes que la definida en HL7, aunque si en un caso determinado algún sistema a incorporar sigue esta norma la ontología puede extenderse en la medida de lo necesario. Del mismo modo los modelos de información relativos a seguridad pueden basarse en el EN13606-4, que trata este tema, y la gestión semántica puede estar inspirada en el EN13606-2, donde se define el uso de arquetipos para la gestión de la historia clínica. Aunque por supuesto todos estos modelos deberán ser refinados para cubrir todas las necesidades de una arquitectura sanitaria compleja.

La vista de información de HISA propone que los modelos de referencia se correspondan con los grupos de actividades identificados en la vista de empresa. En este sentido una extensión de la vista de información viene impuesta por la ya hecha a la vista de empresa en cuanto a la clasificación de actividades. Los modelos de referencia que es necesario especificar están ahora relacionados con las nuevas actividades y por tanto serán:

- Modelo del sujeto de cuidado
- Modelo de información clínica
- Modelo de gestión de actividad
- Modelo de recursos de organización
- Modelo de usuarios, identificación y credenciales
- Modelo de recursos de arquitectura y autorizaciones

- Modelo de auditoría
- Modelo de semántica
- Modelo de comunicaciones

En la norma, antes de comenzar la especificación de cada modelo de referencia se presentan una serie de características y consideraciones comunes a todas, o muchas, de las clases HISA. Algunas de éstas cubren necesidades de determinadas actividades de la vista de empresa por lo que podrían considerarse como parte del modelo de información de las mismas. El problema es que estas actividades son altamente transversales, y por tanto necesitan una serie de consideraciones que afectan a la estructura básica de todas las clases de la vista de información.

Como se indicó en la vista de empresa, la arquitectura debe ser capaz de gestionar la información manejada diariamente en las actividades de usuario, pero también la base de conocimiento que permite manejar la semántica: conceptos, relaciones, vocabularios, elementos y reglas, necesarias para instanciar un elemento concreto de la información operacional requerida en la organización. Para cubrir estas dos necesidades, la vista de información de HISA especifica que para cada modelo de objetos se pueden distinguir:

- **Objetos Operacionales:** estos objetos modelan las entidades implicadas en las actividades diarias dentro de la organización tanto para el tratamiento de pacientes como para el correcto funcionamiento de la misma.
- **Objetos Descriptivos:** especifican la semántica de los objetos operacionales. Modelan las entidades requeridas para el mantenimiento de la base de conocimiento en la que la organización se apoya para ejecutar sus actividades diarias, tanto para el tratamiento a pacientes como para la organización del funcionamiento interno.

Así la relación entre un objeto operacional y un objeto descriptivo puede darse de dos formas, como muestra la figura 27.

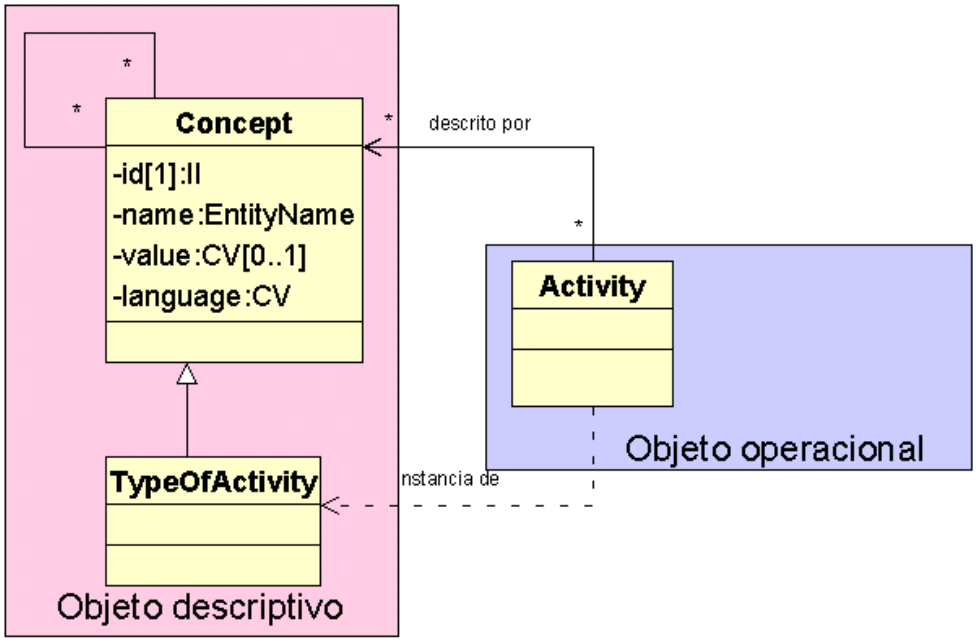


Figura 27. Relación entre objetos descriptivos y operacionales

- Un objeto operacional "es clasificado por" uno descriptivo. Esta relación se representa en HISA con una dependencia que indica que el objeto operacional es una instancia del descriptivo.
- Un objeto operacional "es descrito por" uno o varios descriptivos.

Por otro lado también pueden existir relaciones semánticas entre objetos descriptivos.

Esta forma de modelar la semántica definida en la vista de información es el mecanismo base que HISA provee para la planificación de la gestión de la semántica. Es evidente que el modelo de referencia relativo a las actividades de semántica parte de esta idea, de hecho incluso podría considerarse que el modelo aquí presentado en realidad forma parte del de gestión semántica. El motivo por el que se incluye previamente es que, como ya se señaló en la vista de empresa, al ser las actividades de gestión semántica altamente transversales, es necesario tener presente este mecanismo en todos los modelos de información que soportan las distintas actividades.

En el modelo de información del grupo de actividades de gestión de semántica presentado en esta Tesis, el primero que se va a abordar, se propone un paradigma de diseño que descarta el uso de los objetos descriptivos, pero proporciona una mayor capacidad de expresión e interpretación de semántica, como se verá más adelante.

Todas las clases incluidas en la vista de información tienen la misma estructura: contienen atributos específicos, otros que son comunes a todas las clases y unos atributos extendidos que permiten que en los desarrollos conformes a HISA se puedan incluir los atributos necesarios localmente pero que no se hayan considerado en la norma.

Aparte de estos atributos, como muestra la figura 28, cualquier clase de HISA tiene la posibilidad de incluir reglas relativas a la misma, que pueden ser del tipo que la implementación concreta requiera. Por último se pueden incluir también patrones para poder gestionar el estado de los objetos de información en el pasado. Estos mecanismos cubren dos necesidades:

- La gestión de versiones de un objeto: con el patrón de **`version`**
- La gestión de cambios de estado de un objeto: con el patrón de **`documentation of life cycle`**.

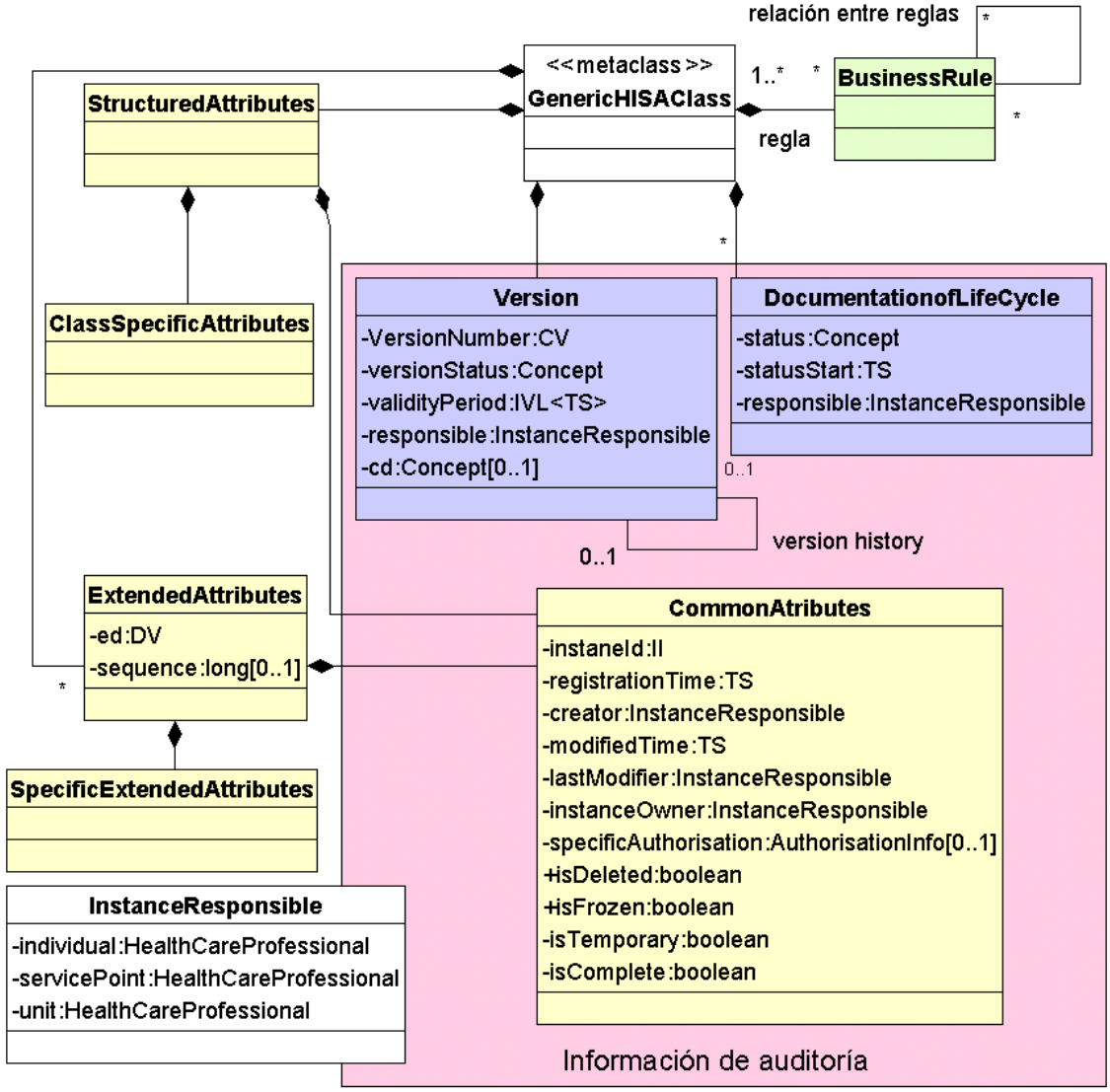


Figura 28. Estructura general de las clases en la vista de información de HISA

Esta estructura general de las clases HISA ya está especificada en la norma. Sin embargo es necesario señalar que tanto los atributos comunes como la información de versión y documentación de estados son mecanismos que van a soportar la realización de las tareas de auditoría. Estas tareas no se incluían en HISA pero se han propuesto como extensión en esta Tesis. Dado que, al igual que las tareas de gestión semántica, las tareas de auditoría son altamente transversales, es adecuado que la información de auditoría vaya adjunta a todas las clases HISA.

También hay que señalar que si se amplía la vista de información con modelos previamente normalizados para cubrir necesidades específicas que HISA en principio no consideraba, será necesario definir procedimientos de consenso entre los atributos de

HISA y los de los modelos cuando éstos se refieren a los mismos conceptos. Así, por ejemplo, en el EN13606 las clases **AUDIT_INFO** y **VERSION** se asocian a determinados objetos dentro de una consulta de historia clínica. Si se utiliza el modelo de información de HISA, algunos de los atributos de estas clases ya están considerados, ya que incluye estos términos como atributos comunes.

Este modelo de clase genérica tiene la ventaja de favorecer la integración según el modelo federado gracias a la existencia, en todos los objetos de la arquitectura, de los campos: **creator**, **instanceOwner** y **lastModifier**, todas del tipo **InstanceResponsible**. Estos campos permiten conocer quién ha creado y modificado cada instancia que se maneje y por tanto puede servir para identificar el dominio de dicha instancia. Es preciso considerar que en una federación, dos instancias distintas pueden representar a la misma entidad real. Los componentes deben tener en cuenta esta posibilidad e incluir mecanismos para resolver la identificación y los posibles conflictos en los valores de atributos.

Otra de las facilidades que los atributos comunes avanzan es la gestión de autorizaciones en la arquitectura. Cada objeto contiene un campo **specificAuthorisation** que identifica las políticas de acceso para ese objeto concreto. En la última versión de la norma se indica que el tipo de este campo tiene que ser todavía investigado; en esta Tesis se propone que sería conveniente utilizar una clase desarrollada dentro del modelo de gestión de seguridad denominada **AuthorisationProfile** que representa una política de acceso, ampliando también la cardinalidad de [0..1] a [0..*].

A continuación la norma presenta los distintos modelos de información para cada grupo de actividades. Las contribuciones de esta Tesis son:

- Presentar la vista de información como una ontología base que va a evolucionar de forma distinta en cada implementación concreta para cubrir sus necesidades específicas de integración. En especial se enlaza el paradigma de modelo dual con la teoría de ontologías para especificar el modelo de información y aportar el dinamismo que éste necesita.
- Realizar extensiones y modificaciones al modelo de información pensando en algunas necesidades de la vista computacional y en el paradigma de federación como modo de interconexión de sistemas.

- Indicar algunas referencias a modelos de información ya existentes que se pueden reutilizar o que al menos es necesario considerar para que tengan cabida en la arquitectura, haciendo especial hincapié en los modelos de información que soportan las aplicaciones de gestión de historia clínica.

6.4.2. Modificación del modelo de semántica

Una de las consideraciones más importantes en una arquitectura para la integración de sistemas es proveer de mecanismos para facilitar la integridad semántica y la gestión del conocimiento. Precisamente los modelos de información son el mecanismo para llegar a acuerdos semánticos, y el modelo para gestión de semántica trata la forma de representación y manejo de estos modelos de información. De hecho la propuesta de este trabajo supone realizar profundos cambios en la metodología de definición de toda la vista de información.

El modelo de información propuesto por HISA como base de la gestión semántica es muy simple, ya que las actividades que lo van a usar se supone que también lo serán. Se define un conjunto de clases descriptivas denominadas **`Tipo de xxx`**, que especifican la semántica en forma de clasificación, y las relaciones entre ellas. Todas las clases **`Tipo de xxx`** heredan de la clase principal **`Concepto`** que tiene como argumento un nombre, un identificador, el idioma codificado y el valor codificado, que es opcional. Como se ha presentado en la introducción, algunas clases de los demás modelos de información son instancias de, o están descritas por, determinada clase **`Tipo de xxx`**. Sobre estas clases enlazadas con alguna **`Tipo de xxx`** se pueden establecer clasificaciones y relaciones, que es lo que permite añadir semántica.

Ya se ha indicado que la gestión semántica es una actividad muy transversal, es decir que se realiza en realidad en todos los grupos de actividades. Por eso el estándar comienza definiendo estas clases descriptivas y operacionales fuera de cualquier modelo de información, porque este mecanismo de expresar semántica debe usarse en todos los modelos. No es de extrañar por tanto que la extensión propuesta en esta Tesis Doctoral sobre el modelo de gestión de semántica se adelante a la primera posición y que sea en realidad un cambio de paradigma de trabajo para definir los modelos de información, afectando por tanto a la definición de todos los demás modelos.

El objetivo de la vista de información debe ser proporcionar una ontología extensible y flexible que incluya los conceptos y relaciones principales del dominio que cubre la

arquitectura sanitaria. El beneficio de diseñar una ontología en lugar de un modelo de información tradicional es que se facilita la gestión del modelo federado, en especial su extensión, y se aumenta la capacidad de expresión semántica. Enlazando con el modelo dual presentado en capítulos anteriores dentro de la ontología federada, algunos conceptos se corresponden con el modelo de referencia y otros son creados realizando restricciones en los mismos, pero todos ellos son clases de la ontología.

La ontología define qué conceptos, o clases, son comunes a todos los componentes de la federación, y qué relaciones existen entre estos conceptos. Una clase tendrá propiedades que podrán ser de dos tipos:

- Propiedades de objetos: expresan una relación entre dos clases.
- Propiedades de tipos de datos: expresan un atributo que puede tener la clase y el tipo del mismo.

Sobre estas propiedades se pueden establecer restricciones que sirven para dar mayor expresividad a las relaciones entre clases. Mientras que en un modelo de información tradicional los vínculos que podían establecerse entre clases eran prácticamente los usados en los modelos de objetos (herencia, composición, agregación…) en una ontología el diseñador puede crear nuevos tipos de relaciones entre clases y restricciones en las mismas, lo que aporta una mayor capacidad de expresión semántica.

El modelo de información sólo define las clases; los objetos concretos serán instancias de las mismas y no forman parten del modelo de información. Del mismo modo, en la ontología está clara la separación entre instancias y clases. Sin embargo muchas herramientas de gestión actuales permiten manejar también las instancias, lo que aporta la ventaja de poder usar de forma integrada razonadores que facilitan la gestión del conocimiento. Éstos componentes pueden llegar a conclusiones a partir del análisis de las instancias y hechos existentes en el sistema y actuar en consecuencia, añadiendo nuevos hechos, instancias o ejecutando alguna otra acción. Por ejemplo, si la propiedad **`tiene adjunto`** que relaciona dos personas tiene como restricción cardinalidad 1, en caso de que exista una instancia de persona, *Isabel*, relacionada con dos personas mediante la propiedad **`tiene adjunto`**, automáticamente se detectará una inconsistencia. O si se ha indicado que la propiedad **`tiene residente`** es la inversa de **`tiene adjunto`** y en la instancia *Isabel* se define la propiedad **`tiene adjunto`** que la relaciona con *Laura*, otra instancia de persona, automáticamente el

razonador decide que *Laura* tiene la propiedad **`tiene residente`** que la relaciona con *Isabel*. Es evidente que en la vista de información de HISA no se incluyen las instancias, pero los componentes encargados de hacer razonamientos, que ya vimos que formaban parte de las actividades de gestión semántica, sí considerarán la ontología, las reglas de inferencia y las instancias. Cuando se consideren todos estos elementos se hablará de una base de conocimiento.

Los modelos de información suelen especificarse como un diagrama de clases UML; más concretamente, todos los que se están usando como referencia en esta Tesis lo están. De manera que sería necesario comprobar si la semántica representada en un diagrama de clases puede expresarse como una ontología, usando un lenguaje de especificación de ontologías. Para ello se ha realizado una experiencia con el modelo de información del openEHR. Para esta experiencia se ha utilizado OWL como lenguaje de ontología y el resultado obtenido ha sido publicado dentro de la colección de ontologías de la Universidad de Standford [143]. Por otro lado, participantes del openEHR han mostrado su intención de usar esta ontología en sus desarrollos. Se ha comprobado también que hay relaciones que pueden expresarse en un lenguaje de ontologías y no en un diagrama de clases. Es de suponer por tanto que en el futuro todos los modelos se expresen como ontologías. Si alguna forma de representación de ontologías logra la popularidad de los diagramas UML llegará a ser el utilizado para representar cualquier dominio. Actualmente el lenguaje más extendido es OWL, aunque existen otros lenguajes incluso más ricos y potentes.

La representación visual de OWL, y de otros lenguajes de ontologías, se realiza con diagramas en los que se representan tripletas de dos conceptos, (o recursos, ya que es una forma de representación heredada de RDF) enlazados por un arco que representa la propiedad que los relaciona; aunque no se reflejan suficientemente bien las restricciones. Una propuesta interesante es que el modelo de información de HISA se defina con OWL en lugar de con el diagrama de clases de UML. Aunque en esta propuesta puede parecer que se están mezclando conceptos de la vista de ingeniería, porque se hace referencia a un lenguaje concreto, no es así, ya que aquí se considera OWL, o algún otro lenguaje de ontologías, como el equivalente a los diagramas de clase UML. En esta Tesis se va a mantener el uso de diagramas de clases, facilitando así la comprensión al lector, dado que no se van a definir relaciones tan complejas que no

puedan expresarse con estos diagramas. Sí se utilizará la representación gráfica de OWL en la vista computacional.

Si el modelo de información se define como una ontología, automáticamente todas las clases se convierten en conceptos, y no es necesario utilizar la clase descriptiva **`Concepto`** ni las clases hijas **`Tipo de xxx`**. Considerando que la capacidad de gestión semántica se proporciona en HISA en aquellas clases que se consideran instancias de alguna clase descriptiva **`Tipo de xxx`** la ventaja de considerar absolutamente todas las clases como conceptos es evidente: todas las clases pueden ser gestionadas semánticamente. De manera que la flexibilidad y el dinamismo que se buscaba en el modelo de información están asegurados.

El modelo de información semántico definía las relaciones entre las clases **`Tipo de xxx;`** eso ya no es necesario ya que las relaciones se establecen directamente entre cualesquiera conceptos del dominio a través de propiedades. ¿Qué debe incluir entonces el modelo de información semántico? Pues aquellos conceptos que se utilicen sólo para las tareas de gestión semántica y que no pertenezcan en concreto a ningún grupo de actividades.

En la figura 29 se representa un modelo resumido de los conceptos de una ontología. Aunque a primera vista pueda parecer complicado, cualquier lenguaje de ontologías incluye estos términos. A partir de este momento cualquier clase definida formalmente en HISA será subclase de la clase **`GenericHISAClass`** y por tanto será una clase de la ontología federada, con la ventaja de que será mucho más sencillo incluir nuevas clases y propiedades así como definir relaciones entre ellas.

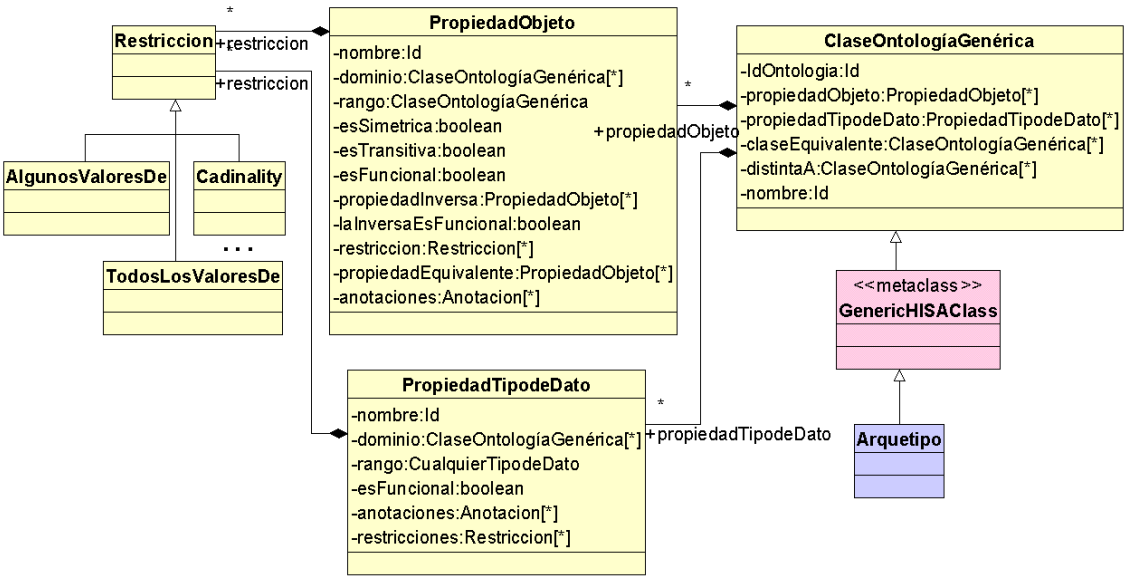


Figura 29. Modelo de conceptos genéricos en ontologías

Las actividades de gestión de ontologías permiten a los usuarios de la arquitectura introducir nuevos conceptos y relaciones. Por supuesto también permiten consultar las ontologías definidas, buscar clases equivalentes a una dada, relaciones de una clase con otras, etc. Todas estas tareas son fundamentales en un modelo federado. Si un nuevo sistema se quiere añadir a la federación necesitará establecer las relaciones de sus conceptos locales con los federados. Las técnicas de gestión de correspondencias entre ontologías permiten buscar estas equivalencias de forma automática o semiautomática, y los lenguajes de ontologías expresarlas. De manera que un modelo ontológico del dominio, en lugar del tradicional modelo de información, facilita la integración de sistemas con distintos modelos.

Si se analizan los modelos de gestión semántica actualmente utilizados en las normas para el desarrollo de sistemas de gestión de historia clínica, tanto en el openEHR como en el EN13606-2 se ha adoptado el concepto de arquetipo. Como ya se comentó anteriormente, un arquetipo permite al usuario definir conceptos, en principio del dominio clínico, a partir de otros conceptos básicos, los del modelo de referencia, y estableciendo restricciones sobre los mismos.

Para que un componente pueda entender un arquetipo es evidente que debe conocer el modelo de referencia en el que está basado. En un modelo federado dos sistemas pueden utilizar distintos modelos de referencia, de manera que es necesario relacionar los dos para encontrar las equivalencias entre ambos y que dos entidades de distintos sistemas,

con distintos modelos, puedan tener una semántica común. Si los modelos de referencia y arquetipos se representan como ontologías se facilita la expresión de estas relaciones.

Un arquetipo podría verse como una clase de la ontología que representa un concepto concreto modelado directamente por el usuario estableciendo restricciones sobre otras clases. De manera que en realidad la idea es muy similar salvo que mientras que los arquetipos han sido específicamente pensados para el dominio clínico, y por tanto son especialmente aptos para representar la semántica de este tipo de información, las ontologías son de propósito general. Normalizar HISA en términos de ontologías permitiría una mejor reutilización de componentes y una mayor integración con otros dominios que pueden estar implicados, hoy o en el futuro, en una organización sanitaria. Para ser totalmente compatibles con el EN13606-2, y por supuesto facilitar la incorporación a la arquitectura a todos los sistemas que utilicen la técnica de arquetipos para representar semántica, es conveniente introducir el concepto **`Arquetipo`**, tal y como lo define la norma, como una clase más de la ontología, como muestra la figura 29 de modo simplificado.

Otro concepto importante en la gestión semántica es la definición de reglas. Estas reglas, junto a la definición de la ontología y las instancias de la misma, servirán para:

- Hacer razonamientos generando nuevo conocimiento que añadir al sistema.
- Disparar acciones, de manera que pueden ser el mecanismo para el comportamiento proactivo de algunos componentes.

Una regla para inferir nuevo conocimiento sigue el modelo:

`Si Antecedente Implica Consecuencia`.

Si en la base de conocimiento se cumplen las condiciones indicadas en el **`antecedente`** entonces también se cumplen las especificadas en la **`consecuencia`**. En el **`antecedente`** se consideran las características conocidas sobre las instancias de la base de conocimiento y en la **`consecuencia`** se indicará una nueva afirmación que se puede realizar sobre éstas.

Una regla que dispare acciones sigue el modelo:

`Si Condición realizar Acción`.

Pero no se ejecuta en un instante predeterminado ni en un orden específico, como ocurre en programación. La **acción** se ejecuta en el instante en el que la **condición** cambie de FALSA a VERDADERA. Esta **condición** consiste en un conjunto de patrones que se comparan con los hechos de la base de conocimiento (o pizarra): en este caso sería la base de conocimiento de la arquitectura sanitaria o una parte de ella. La **acción** es la ejecución de algo programado, como por ejemplo la invocación de una operación o el envío de una señal.

Como se puede apreciar, un **antecedente** y una **condición** son prácticamente lo mismo, mientras que una **consecuencia** implica un cambio de la base de conocimiento y sin embargo una **acción** indica la ejecución de alguna operación, que por supuesto puede introducir algún cambio en la base de conocimiento.

En la figura 30 se presenta un modelo simplificado de cómo se podrían expresar reglas de inferencia y de acción. Este modelo simplificado está basado en el lenguaje SWRL [93] y en el lenguaje de expresión de reglas de Jess [144]. Como se puede apreciar, la conexión con el resto de las clases de la vista de información es a través de los átomos, donde se expresan afirmaciones sobre clases o propiedades del modelo. A continuación se presenta un ejemplo de cómo se podrían manejar reglas para la gestión de personal en un hospital.

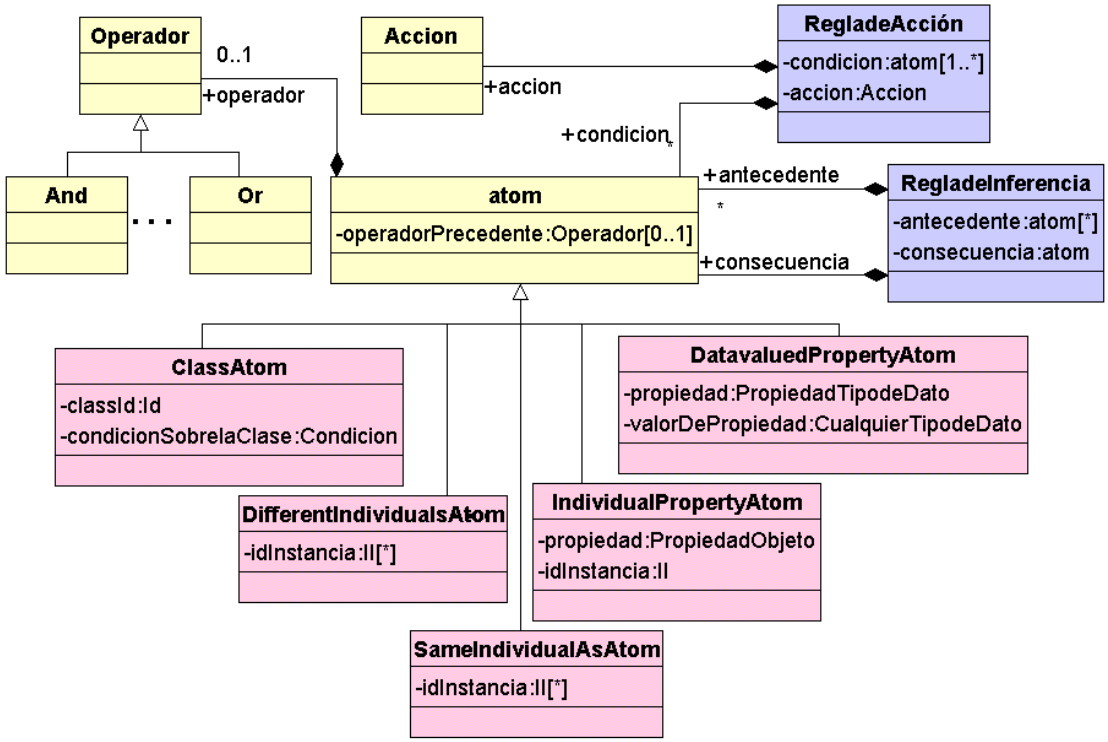


Figura 30. Modelo de reglas

Durante las vacaciones de la *Doctora Laura* todos los residentes a su cargo (relacionados con ella mediante la propiedad **`tiene adjunto`**) van a pasar a depender de su colega *Germán*, para ello se podría crear una regla de inferencia que tuviera la siguiente estructura:

Como antecedente un átomo del tipo **`IndividualPropertyAtom`**, con la propiedad tiene adjunto y el identificador de instancia *Laura*. Para todas las personas cuyo adjunto fuera *Laura* sería cierto este átomo, por ejemplo *Isabel* cumpliría la condición.

La consecuencia de la regla sería un átomo del tipo **`IndividualPropertyAtom`** en el que la propiedad fuera **`tiene adjunto`** y el identificador de instancia *Germán*. Como *Isabel* cumplía el antecedente debe aplicarse la consecuencia, así que a partir de ese momento la propiedad **`tiene adjunto`** pasa a relacionarla con *Germán* y no con *Laura*.

Las reglas de acción estarían más relacionadas con el comportamiento de los componentes, por lo que para la especificación de la vista computacional y, sobre todo, de las de más bajo nivel, se puede recurrir al uso de estas reglas que han sido previamente modeladas en la vista de información.

Las actividades de gestión de conocimiento están muy relacionadas con los razonadores, que básicamente ejecutan las reglas definidas, y pueden ser el soporte, por ejemplo, de aplicaciones de apoyo a la decisión clínica.

En cuanto a la gestión de anotaciones, en esta Tesis se define una anotación como un elemento de información ligado a algún otro elemento de información de la arquitectura y cuya finalidad es matizarlo, extenderlo o relacionarlo con otro. Hay que tener en cuenta que la anotación puede cambiar notablemente el significado de una instancia. Las anotaciones pueden ser totalmente externas al elemento de información al que se refiere o estar incluidas dentro del propio elemento de información. Este segundo caso es especialmente interesante en las anotaciones sobre información en formato texto, muy habitual en la práctica diaria de la medicina. Puede resultar útil realizar las anotaciones dentro del propio texto, por ejemplo usando XML y las etiquetas definidas en algún esquema que dé significado a las mismas, aunque por motivos de seguridad podría ser necesario mantener además la versión original.

La figura 31 presenta un modelo muy simplificado de anotaciones, donde se han identificado sólo dos tipos básicos, pero pueden extenderse. Todas las anotaciones están

referidas a alguna instancia dentro de la arquitectura. El significado de la **AnotaciónCodificada** se da a través de un código, dado que este código se considera del tipo **CV (Coded Value)** del CEN y que éste tipo incluye también el sistema de codificación utilizado, se facilita la interpretación correcta de la anotación. Por ejemplo la anotación de una observación clínica con el código **DS03** según el EN13606-3 indica que el sujeto al que se refiere la información de esa instancia es la madre del paciente. Si una anotación codificada relaciona dos instancias a través de un código sería de tipo **AnotaciónCodificadaRelación**. Por ejemplo la anotación de la persona *Carmen* con el código **DS03** según el EN13606-3 y en relación con *Isabel* indica que *Carmen* es la madre de *Isabel*.

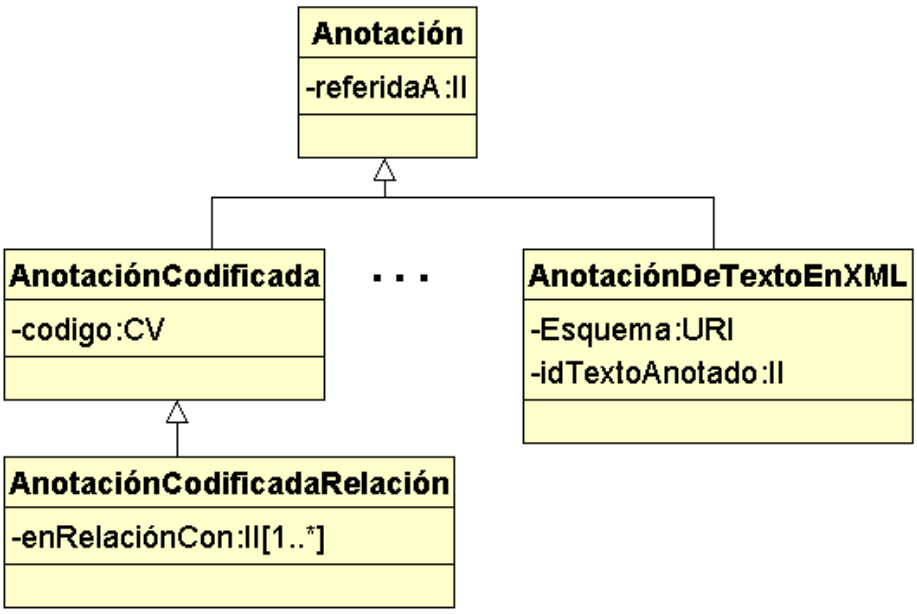


Figura 31. Modelo simplificado para anotaciones

Precisamente la necesidad de expresar el significado de conceptos como combinación de otros conceptos representa una de las mayores dificultades en la gestión semántica. Un buen modelo de anotaciones ayuda a soportar estas tareas.

En cuanto a las anotaciones de texto con XML sólo destacar que es necesario indicar el esquema que se está utilizando, para poder interpretar las etiquetas utilizadas, y que en el caso de que el texto anotado reemplace al original es necesario seguir los mecanismos de gestión de versiones impuestos por la arquitectura.

Las anotaciones pueden ser hechas en el momento en el que se crea la información o posteriormente. Como la anotación es una clase HISA, datos como el instante de creación, el responsable, u otros de interés para las tareas de auditoría, están perfectamente registrados. En HISA no se incluían clases para gestionar estas anotaciones y en los estándares que se están utilizando como referencia los mecanismos son muy básicos. Tanto en EN13606, a través de las anotaciones incluidas en la clase **`ENTRY`**, como en COAS, con los **`Calificadores de observación`**, sólo se puede anotar la información clínica. En ambos casos los tipos de anotaciones están definidos en un sistema de códigos, en el primero normalizado en la tercera parte del estándar y en el segundo no se especifica, siendo por tanto dependiente de la implementación.

Para terminar sólo indicar que este modelo seguirá creciendo a medida que las necesidades de gestión semántica crezcan y que la principal aportación es considerar el modelo de información como una ontología. Esta consideración proporciona un dinamismo intrínseco del que carecen los tradicionales modelos cerrados, que únicamente pueden crecer según los mecanismos de extensión previstos en los mismos, que suelen ser bastante limitados.

6.4.3. Extensión del modelo del sujeto de cuidado

Ya que otra de las aportaciones de esta Tesis Doctoral es precisamente el modelado de componentes para la identificación de personas y la gestión de sus datos demográficos, este modelo de referencia ha sido especialmente extendido para cubrir todas las necesidades que surgen para llevar a cabo estas actividades en un entorno federado. Estas extensiones son presentadas más adelante en detalle, en este punto sólo se muestra una visión a un alto nivel de abstracción. Los modelos de información que se han considerado para realizar las extensiones son los GPIC del CEN, en particular los que se refieren a datos demográficos, y el modelo del servicio PID de CORBAmed.

El concepto fundamental del modelo de información del sujeto de cuidado es la persona. En la clase **`Person`** de HISA se incluyen los atributos que tiene toda persona. Los tipos de estos atributos son tipos normalizados por el CEN. Si todos los sistemas de una federación modelaran la persona con la misma clase el problema semántico estaría resuelto, pero es evidente que cada sistema local utiliza su propio modelo de persona y que éste tiene que ser convertido al federado. Un conjunto cerrado de atributos no es la

mejor opción para la clase **`Person`** en la ontología federada ya que el establecimiento de equivalencias desde el sistema local al común se hace más complejo. Es mucho más eficiente que el conjunto de atributos de la persona sea abierto, expresable por los usuarios mediante la ontología de alto nivel, aunque se indiquen algunos como obligatorios. En ese caso el interés es poner a toda la federación de acuerdo en la semántica de los distintos atributos y ser capaces de establecer relaciones entre ellos y mecanismos que faciliten las correspondencias de atributos locales a federados. Desde este punto de vista la forma en que el modelo de información del servicio PID trata los atributos de persona, modelo que vimos en el capítulo de servidores demográficos, es mucho más eficiente. Hay que tener en cuenta que si los componentes de gestión demográfica son suficientemente flexibles, permitirán ser interrogados para solicitar atributos específicos de una persona y poder crear un objeto persona contextualizado. Así, un objeto persona puede ser distinto en distintos escenarios y para distintos fines, pero nunca habrá problemas de interpretación semántica.

Se propone por tanto utilizar el modelo de información del PID para modelar los atributos de persona y la ontología de alto nivel, por ejemplo basada en los GPICS del CEN, para particularizar la semántica compartida de estos atributos. Esto asegura que todos los atributos considerados en el modelo de información de HISA siguen estando, pero se extiende el número de atributos a la vez que se facilita la futura extensión y los mecanismos de establecimiento de correspondencias. Para especificar los atributos concretos que se incluyen en un objeto persona, que como se ha indicado ahora pueden ser variables, así como las restricciones impuestas a éstos, y para seguir la misma filosofía del EN13606, se puede recurrir al concepto de arquetipo demográfico. Estos arquetipos estarán gestionados a través de las actividades semánticas.

Por otro lado, como ya se ha señalado anteriormente, dado que cada sistema suele tener su propio mecanismo de registro de personas, la información demográfica acerca de una persona puede estar muy duplicada en una federación. Además será muy usual que la misma persona esté registrada en varios sistemas pero que algunos de los atributos almacenados no coincidan. En este sentido el modelo de información debe facilitar también la resolución de conflictos de entidad, como se verá más adelante. Un modelo de información adecuado puede facilitar a los componentes, en caso de que fuera necesario, la creación de una instancia de persona que sea válida en toda la federación y

que se corresponderá con una instancia de la misma persona, la entidad real, en cada uno de los sistemas, como se muestra en la figura 32.

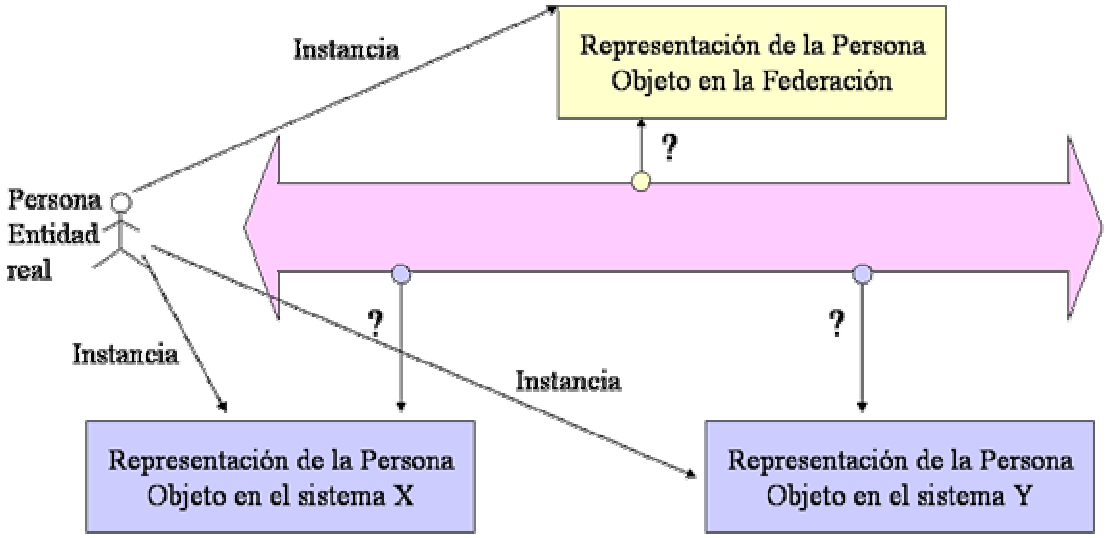


Figura 32. El objeto persona en un modelo federado

Conceptos considerados en el modelo de HISA y relacionados con la persona son el **`SujetoDeCuidado`**, que representa a un paciente, y el **`ProfesionalSanitario`**, entre otros. Dado que estas clases pueden contener nuevos atributos distintos de los contenidos en la clase genérica **`Person`** el mecanismo de especificación de estos atributos debe ser semejante al de persona, para facilitar la compatibilidad semántica. Sería conveniente introducir en el modelo el concepto de **`Rol`**, ya que una misma persona puede ser sujeto de cuidado y profesional sanitario dependiendo del contexto y actividad concreta. La forma de representar roles y relaciones en los GPICS puede ser la base de esta extensión.

También son conceptos considerados en este modelo del sujeto de cuidado el **`Contacto`** y el **`PeriodoDeCuidado`**. Un contacto es una situación en la que un proveedor de servicios sanitarios realiza algún servicio para un paciente y/o accede a la información de su historia clínica para consultarla y actualizarla. El periodo de cuidado es una estructura que agrupa las condiciones de salud y/o contactos y asocia elementos que están almacenados en el sistema de información y que conciernen al sujeto de cuidado en el contexto de una condición sanitaria concreta. Estas dos clases permiten enlazar el modelo de información del sujeto de cuidado con el de gestión de actividades y con el de información clínica. Cada contacto estará en el marco de un periodo de

cuidado y se refiere a un paciente, de modo que en cierta actividad de la organización se pueden realizar contactos que provocan la consulta y creación de información clínica relativa a ese paciente.

Otro modelo demográfico que se debería tener en cuenta en un estándar europeo es el utilizado en el EN13606, más concretamente el definido en la parte 1. En realidad los datos demográficos de este modelo son los descritos en los GPICS, así que no hay ningún tipo de incompatibilidad con el de HISA ni con la extensión propuesta en esta Tesis. Por otro lado en este modelo no existen los conceptos de contacto, periodo de cuidado y actividad, aunque sí el de sesión clínica. Esto nos da una idea de que el estándar 13606 está demasiado centrado en la historia clínica y no considera algunos conceptos importantes que hay que tener en una arquitectura de sistemas sanitarios de mayor cobertura.

Para terminar, recordar que este modelo se extenderá en el próximo capítulo de esta Tesis Doctoral, en el que se presentarán las vistas de empresa, información y computacional específicas de la gestión de información demográfica.

6.4.4. Extensión del modelo de información clínica

En cuanto al modelo de información clínica, es necesario considerar que la información clínica tiene naturalezas muy diversas; imágenes, resultados de laboratorio, etc, y de hecho el tipo de información que más abunda es la información de texto libre. Se puede observar que la propuesta de HISA es utilizar la clase **`InformaciónClínica`** para representar cualquier dato clínico. Este modelo es muy sencillo y no cubre toda la semántica necesaria para especificar la información clínica. Por otro lado el único mecanismo de extensión soportado es a través de la clase descriptiva **`TipodeInformaciónClínica`**.

Una ontología que contenga más semántica para la representación de observaciones clínicas podría facilitar las tareas de recuperación, modificación y almacenamiento de este tipo de información y mejorar las condiciones de gestión de este conocimiento en entornos federados. Además se facilita la extensión y la correspondencia de términos desde distintos dominios. En ese sentido, en el servicio COAS de CORBAmed y en el EN13606-1 se proponen modelos de información más específicos y completos que podrían ser la base de la ontología de información clínica de HISA. Así, considerar la clase **`InformaciónClínica`** equivalente al concepto de **`ObservaciónClínica`**

del COAS o al concepto **`Item`** del EN13606 proporcionaría un marco más flexible a la hora de representar información clínica. La figura 33 muestra, de forma simplificada, estos dos modelos y las equivalencias más sencillas que se pueden observar entre ambos.

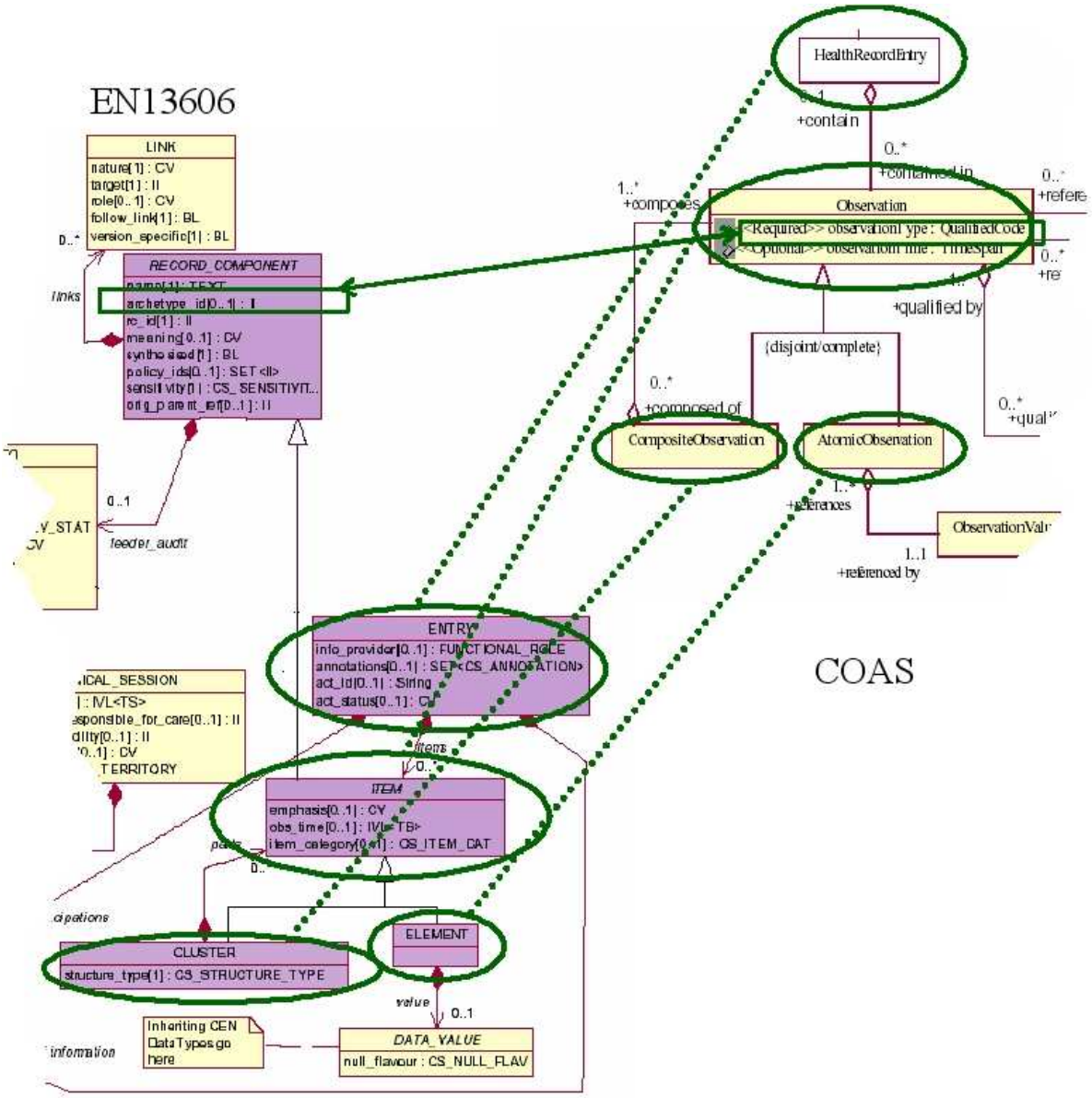


Figura 33. Relación entre los modelos de información del COAS y del EN13606 (Simplificado)

Estos modelos sólo dan los contenedores y el mecanismo de organización de las observaciones clínicas, de manera que la semántica concreta sobre qué almacena cada observación tiene que ser también añadida. El mecanismo para especificar semántica en COAS es utilizar el campo **`TipodeObservación`**, sin embargo en el EN13606 se

utiliza el concepto de **arquetipo**, ambos marcados en la figura 33. El mecanismo de arquetipos es mucho más flexible y el manejo de éstos, como ya se ha visto, se facilitará a través de las actividades de gestión semántica. Por tanto en una primera aproximación parece más adecuado extender el modelo de información clínica con los conceptos manejados en el EN13606.

Volviendo a los dos modelos que nos ocupan, es especialmente importante observar la relación entre la clase **CalificadordeObservación de COAS** y el parámetro **annotations** de la clase **ENTRY** en el EN13606: tienen un uso muy similar y equivalente a las anotaciones definidas en esta Tesis Doctoral.

Aunque parece que presentar el modelo COAS aquí no es necesario, se ha querido mostrar por dos motivos. Primero porque es interesante observar las equivalencias entre dos modelos muy aceptados en la comunidad científica. Será muy habitual en una arquitectura de esta naturaleza realizar este tipo de correspondencias entre distintos modelos semánticos. Utilizar ontologías para la descripción del dominio, como ya se ha indicado, va a facilitar estas tareas. También es importante presentar COAS porque en esta norma se define un modelo de interfaz para recuperar observaciones clínicas, cosa que por ahora no se hace en el EN13606, y que puede ser utilizado como base para la vista computacional de HISA, con lo que es interesante tener presente el modelo de información del que parte y las relaciones entre éste y la ontología federada que, en este caso, se propone esté basada en el EN13606-1.

Se necesita realizar mucho más trabajo para definir la semántica exacta de cada tipo de observación; este esfuerzo puede facilitarse si se adoptan estándares para la representación de tipos concretos de observaciones, de modo que HISA debe dar una orientación en este sentido. Por ejemplo, para representar objetos que almacenen imágenes puede utilizarse el modelo de datos propuesto en el estándar DICOM [49], para los análisis de laboratorio pueden utilizarse los modelos de LOINC [64,145], etc. Aunque siempre hay que tener en cuenta que el objetivo es facilitar la gestión de cualquier tipo de información.

Es evidente que aunque se utilice el modelo EN13606 para representar la información clínica, las actividades de auditoría y control de versiones deben ser comunes al resto de los modelos de información, de manera que es necesario establecer las equivalencias

entre los campos de la clase **`Audit_info`** y **`Version`** de este modelo y los atributos comunes, la clase **`Version`** y **`DocumentationofLifecicle`** de HISA.

Gracias al modelo semántico que hemos elegido, y a la consideración de todo el modelo como una ontología, todas estas equivalencias de clases y atributos y las relaciones nuevas que se establecen pueden ser fácilmente introducidas en la ontología federada. Del mismo modo se pueden incluir las clases de COAS dentro de la ontología y definir las relaciones entre los dos modelos para facilitar la integración de sistemas basados en COAS.

Otro hecho a considerar es que tanto en el modelo propuesto por HISA como en el EN13606-1 y en COAS se considera que la información clínica siempre está referida a un sujeto de cuidado determinado. Sin embargo, en una arquitectura que cubra otras necesidades de la organización sanitaria puede necesitarse acceder a la información clínica sin considerarla como parte de la historia clínica de un paciente, de manera que estas observaciones puedan ser "despersonificadas", es decir que pueda perderse, o al menos no mostrarse, la relación de una observación con un paciente determinado.

Otro de los problemas en un entorno federado es la reconciliación de instancias; es decir: identificar si dos instancias se refieren a la misma entidad real y si es así resolver los posibles conflictos en sus valores. La información clínica suele estar menos duplicada que la demográfica y cuando lo está generalmente es por la copia en un sistema de la información originada en otro, al contrario que la información demográfica que generalmente siempre tiene un origen local a cada sistema. Esto quiere decir que los conflictos de instancias van a ser menores y que muchos de ellos pueden resolverse únicamente con la información de auditoría, aunque es necesario considerar la posibilidad de casos más complejos.

Por último señalar que aunque según todos estos comentarios da la impresión de que el modelo de información clínica de HISA es deficiente, en realidad una de las aportaciones más importantes de este modelo es especificar los enlaces entre las clases del modelo de información clínica y las del resto de modelos de la vista de información. Estas relaciones tienen que seguir manteniéndose de manera que la única modificación propuesta es sustituir la clase simple **`ClinicalInformation`** por un conjunto de clases que representa el mismo concepto pero de forma más compleja, permitiendo así mayor flexibilidad.

6.4.5. El modelo de gestión de actividad

En este modelo se incluyen todas las clases que van a dar soporte a las aplicaciones de gestión de las actividades que tienen lugar en la organización sanitaria y al seguimiento de los ciclos de vida de las mismas según la planificación realizada. También se consideran las actividades de gestión de guías clínicas y planes relativos a una condición sanitaria específica o a sujetos de cuidado particulares.

El estándar presenta dos modelos de información que cubren estos dos grupos de actividades, que son las clases principales del modelo, y la gestión de semántica y clasificación de las mismas se hace a través de las clases descriptivas **`TipodeActividad`** y **`TipodePlan Clínico`**, mecanismo habitual de HISA. Esta separación sólo es para facilitar la comprensión y en realidad están enlazados a través de las clases **`Actividad`** y **`PlanClínico`**.

Dentro del EN13606 no se han incluido elementos de información que soporten la gestión de actividades y de guías y planes al no considerarse estrictamente parte de la historia clínica de una persona. Sin embargo es importante destacar que una actividad puede estar relacionada con un determinado sujeto de cuidado e incluir contactos entre éste y la organización sanitaria, de manera que necesariamente existen nexos entre los modelos de información de gestión de actividad y de sujeto de cuidado.

No se presenta ninguna extensión a este modelo de información, aunque se aventura que será un modelo muy dinámico, de manera que las extensiones sobre el mismo deben ser realizadas a través de las actividades de gestión semántica, modificando la ontología federada.

6.4.6. Modelo de recursos de organización

No se propone ninguna extensión a este modelo. La clase fundamental es **`Recurso`**, cuya semántica se especifica a través de la clase descriptiva **`TipodeRecurso`** pero que el modelo ontológico propuesto permitirá una mayor expresividad de los mismos.

Como se indicó en la vista de empresa, los recursos de la organización se refieren a todos aquellos medios que se utilizan en la organización sanitaria y sobre los que es necesario establecer ciertas políticas de uso. Los recursos implicados en una organización sanitaria son prácticamente innumerables pero se podría hacer una pequeña clasificación en:

- **Recursos humanos:** el personal de la organización, entre el que encontraremos responsables de tareas administrativas, de asistencia sanitaria, limpieza, lavandería, restauración o financieras entre otras muchas.
- **Material consumible:** Incluye todo lo que se puede utilizar tanto en la actividad asistencial como en las tareas administrativas. Al ser material fungible es necesario tener controlado el número de unidades disponibles, la localización de las mismas, la necesidad prevista y el gasto.
- **Dispositivos y Localizaciones:** Se incluye en este grupo cualquier recurso limitado que no sea consumible. Entre estos se consideran las camas de hospital, las consultas, los quirófanos o la maquinaria para determinada terapia o diagnóstico

Es importante señalar que los recursos pueden ser utilizados dentro de determinada actividad y que por tanto la disponibilidad de estos recursos puede influir en el estado de estas actividades sanitarias y de los contactos implicados en ellas. El modelo de información debe incluir todas estas relaciones y la gestión deberá hacerse a través de las actividades de gestión de recursos de la organización.

6.4.7. Modelos de Seguridad y Gestión de recursos

Modelo para la gestión de usuarios, identificación y credenciales

En este caso la ontología definida tiene que dar soporte a las actividades relacionadas con la seguridad en términos de gestión de usuarios del sistema, la correcta identificación de los mismos y la obtención de sus credenciales.

En el modelo de HISA se define usuario como la entidad lógica que representa a un profesional sanitario concreto o a una funcionalidad del sistema, es decir un componente, que quiere acceder a un recurso de la arquitectura de acuerdo a unas autorizaciones dadas. El primer cambio propuesto respecto al modelo es que también se considera usuario cualquier persona que quiera utilizar las aplicaciones que soportan las actividades identificadas en la vista de empresa. Así, un usuario, además de un profesional que pertenece a la organización sanitaria, puede ser un sujeto de cuidado o alguien totalmente ajeno al sistema. Se considera importante hacer esta matización ya que es necesario tener en cuenta que la arquitectura puede proveer servicios que no estén orientados al profesional de la organización sino a los pacientes o familiares, a

profesionales de otros entornos, o al público en general. De manera que estos tipos de usuarios deben considerarse también en el modelo. Ya que un usuario puede ser una persona, es evidente que el modelo de **`Persona`** escogido en HISA influirá en el modelo de **`Usuario`** y que muchos de los componentes que se consideran del grupo de actividades de gestión del sujeto de cuidado pueden reutilizarse para estas actividades de seguridad.

En una primera aproximación se pueden distinguir dos grandes grupos de usuarios: los que pertenecen al sistema y los que no. Si el usuario no pertenece al sistema, las tareas se limitarán, en el caso más complejo, a realizar identificaciones fallidas o, en caso de acceder a facilidades libres, registrar al usuario solicitando información útil al sistema y que servirá básicamente para las tareas de auditoría. Evidentemente la confianza en los datos introducidos por el usuario es baja a menos que exista algún mecanismo añadido para la verificación.

En cuanto a los usuarios que pertenecen al sistema, las tareas de alto nivel implican el registro del usuario y la gestión de sus credenciales, así como la identificación y autenticación cuando se acceda a los recursos de la arquitectura. En un modelo federado la complejidad aumenta, ya que el usuario puede estar registrado en cualquiera de los dominios que pertenecen al sistema, o en varios de ellos, y tanto la identificación como las credenciales pueden ser distintas para cada uno. La ontología que sirva de apoyo a los componentes que lleven a cabo acciones para completar estas tareas de seguridad debe dar cobertura a todas estas consideraciones.

En el EN13606-4 se normalizan algunos conceptos relativos a la seguridad en los sistemas de historia clínica; algunos de éstos concuerdan con la idea de gestión de usuarios y credenciales y pueden ser utilizados para aumentar la ontología de HISA en esta materia. En la EN13606-4 se considera que los privilegios son asignados a roles a través de certificados de especificación de rol. Por otro lado los roles son asignados a los usuarios a través de certificados de asignación de roles, que pueden contener uno o más roles. Se podría considerar que un certificado de este tipo es lo que estamos denominando credencial de usuario. En HISA, la clase más cercana a este concepto es **`AuthorisationRole`**. Incluyendo las ideas del EN13606-4, más la gestión de identificación y la utilización de identificadores de dominio para mejorar las condiciones de gestión en entornos federados, el modelo quedaría como muestra la figura 34.

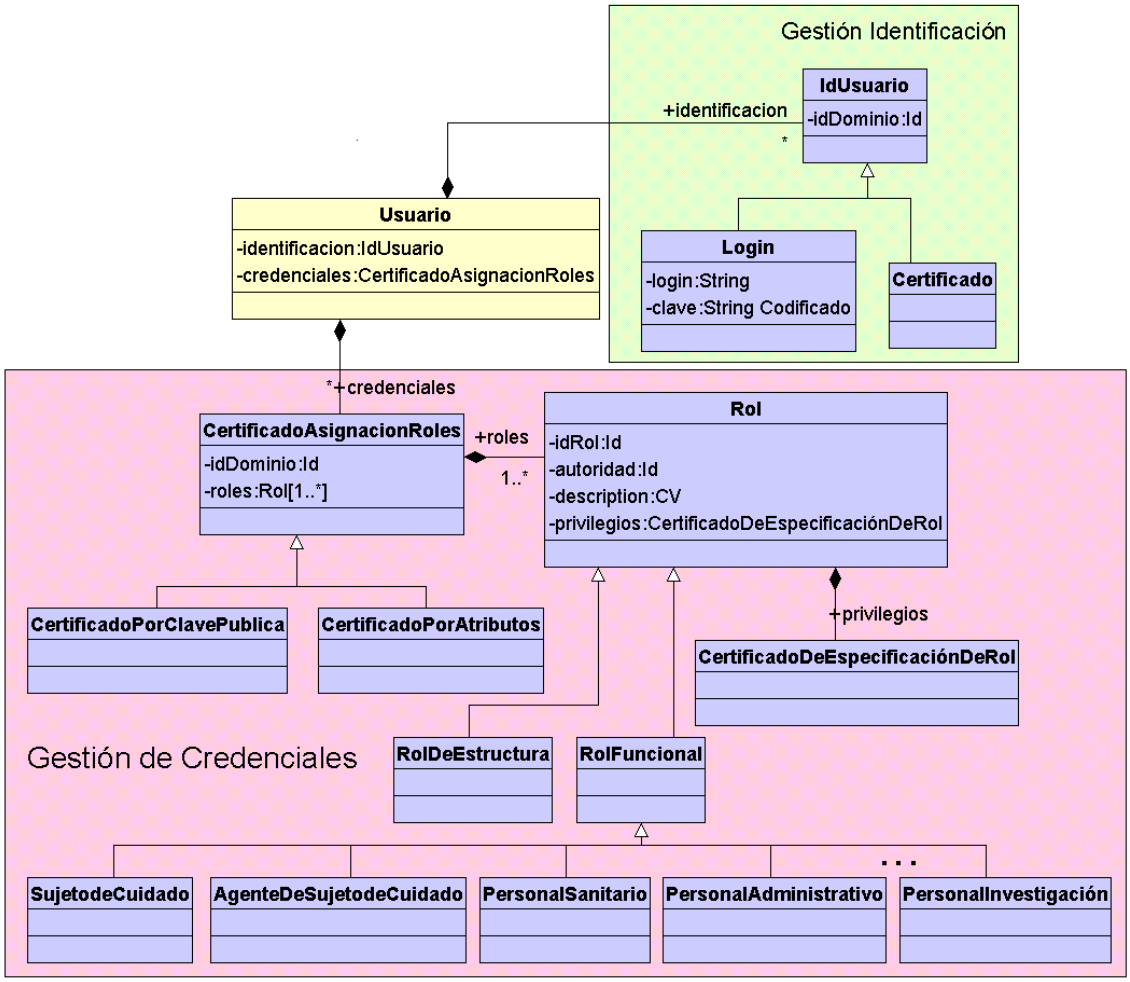


Figura 34. Modelo para la Gestión de usuarios, identificación y credenciales

Respecto a las modificaciones del modelo original de HISA, hay que destacar que la identificación de usuario se ha sacado fuera del usuario para considerar otros casos distintos al uso de usuario/clave. Es importante que el modelo contemple esto entre otras cosas porque, como ya se ha dicho, un usuario puede ser un componente, que difícilmente se identificará con el mecanismo usuario/clave. En el modelo propuesto se incluye también el caso de utilización de certificados, pero podría extenderse creando nuevas subclases de **`IdUsuario`**. Por otro lado el uso de certificados digitales cada vez está más extendido como mecanismo de identificación de usuarios.

En este modelo, sólo se consideran las credenciales que en HISA se expresaban en la clase **`AuthorisationRole`** equivalente a la clase **`CertificadoAsignacionRoles`** de la figura 34. No se consideran las autorizaciones, ya que dan soporte a actividades de control de acceso que en la modificación propuesta están separadas de las de identificación y gestión de usuarios y

credenciales. Aunque ambos modelos estarán muy relacionados, dada la clara especialización de las actividades y la complejidad de las mismas se ha considerado mejor considerarlas por separado. Tampoco se consideran los recursos, que están también recogidos en el modelo que a continuación se describe.

Modelo de recursos de arquitectura y autorizaciones

En cuanto a las tareas relativas a la gestión de recursos de la arquitectura y las autorizaciones de uso de las mismas, es evidente que estarán muy relacionadas con las anteriores, ya que las políticas de acceso se definirán en términos de las credenciales de la entidad que intenta acceder. Dos conceptos son fundamentales para llevar a cabo estas tareas: el recurso y las políticas de acceso al recurso.

Un recurso es cualquier elemento perteneciente a la arquitectura. Se pueden identificar dos tipos básicos de recursos: elementos o piezas de información y componentes o funcionalidades. En el modelo de información para gestión de autorizaciones de acceso HISA define las clases **`ControlledInformationElement`** y **`ControlledFunctionality`** que se corresponden con esta idea.

Una política de acceso a un recurso define qué características tiene que cumplir la entidad que accede a ese recurso y qué tipo de acciones puede realizar sobre el mismo. Para un recurso se puede definir varias políticas de acceso y habrá que decidir qué políticas tienen que ser evaluadas en un instante determinado. En el modelo HISA, el concepto que recoge esta idea es la clase **`AuthorisationProfile`**, que es una regla de negocio del tipo **`BRConsent`** que sería el equivalente a política de acceso, y que relaciona los recursos con los usuarios a través de la clase **`AuthorisationRole`**. El modelo es muy simplificado y no especifica cómo se definen políticas ni qué tipos de políticas puede haber. Hay que recordar que cada clase genérica HISA contenía un atributo denominado **`specificAuthorisation`** donde el estándar propone que se identifique la regla de autorización del objeto para cualquier usuario. El tipo no estaba decidido aún, y en esta Tesis se propone que sea **`AuthorisationProfile`**, es decir, una política de acceso. También se propone la modificación de la cardinalidad ya que un mismo recurso puede estar controlado por más de una política.

Al igual que en el caso anterior, el EN13606-4 puede servir de referencia para la extensión de la ontología de HISA, y es importante considerar también el modelo

especificado en el servicio RAD de CORBAmed. En el primero se da un modelo de información para políticas, mostrado en la figura 35, y plantea un modelo funcional muy sencillo de acceso a recursos previa autorización. El segundo no especifica cómo es una política, porque queda fuera del servicio de decisión de acceso a recursos, pero sí un modelo de información, mostrado en la figura 36, que incluye algunos elementos útiles a la hora de definir los componentes que forman parte de este servicio y las interacciones entre ellos. Por otro lado especifica un conjunto de interfaces para realizar tareas y que puede servir de base para el desarrollo de la vista computacional HISA.

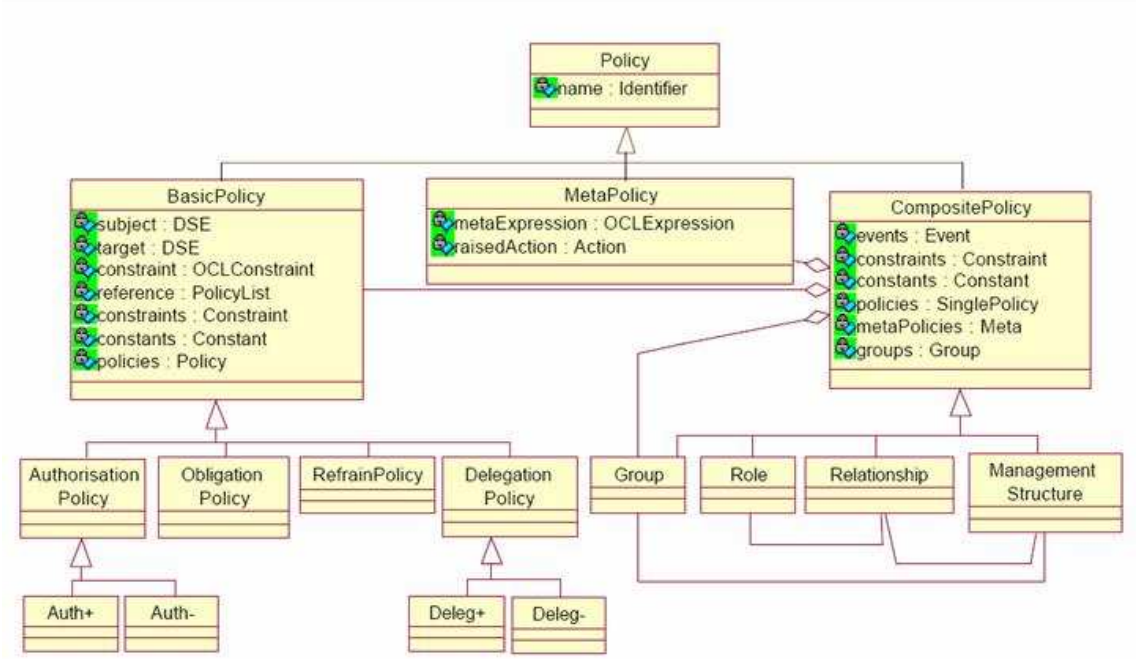


Figura 35. Modelo de información de políticas de acceso del EN13606-4

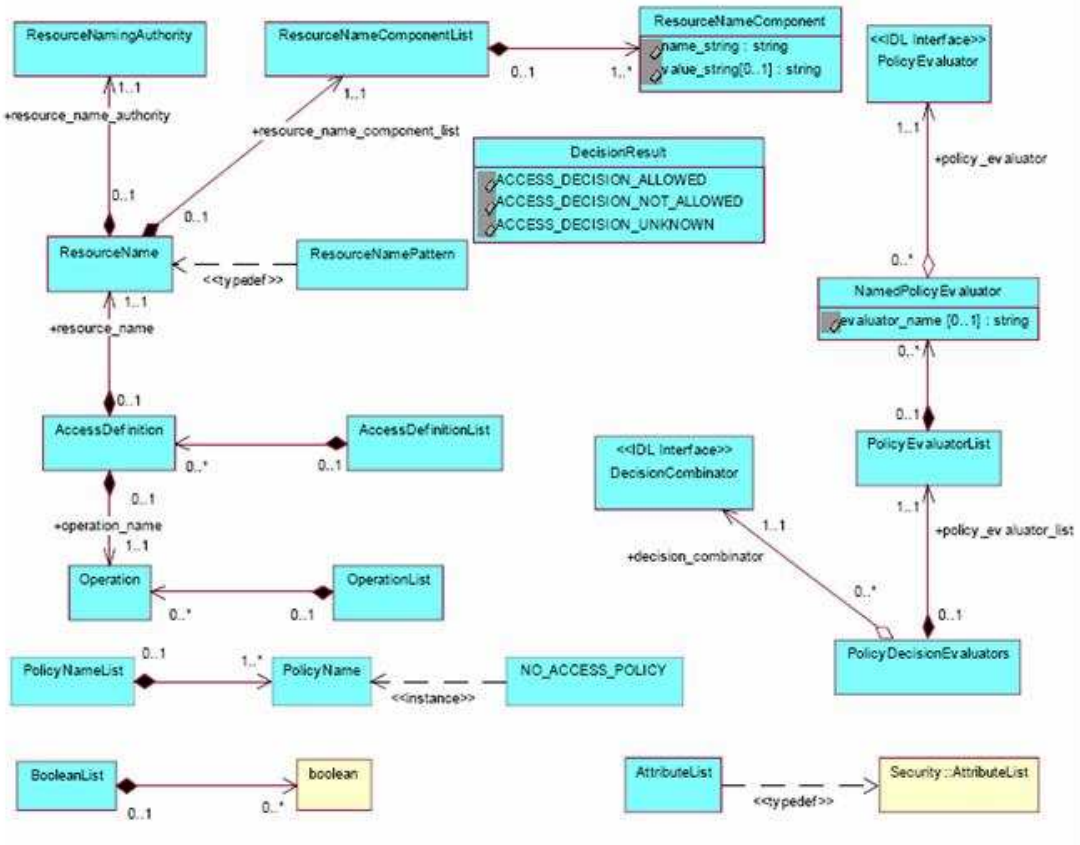


Figura 36. Modelo de información del RAD de CORBAmed

Modelo de auditoría

De un modo muy simplificado se podrían definir las tareas de auditoría como todas aquellas que permiten conocer la historia de lo acontecido en el sistema de información. Los hechos que se deben registrar vienen determinados por las necesidades de la organización sanitaria y por los requisitos legales exigidos a ésta. HISA no contempla estas tareas como un grupo aparte de actividades, sin embargo la estructura de la clase genérica HISA, como se vio en la figura 28, sí que provee de algunos elementos que pueden ayudar al desarrollo de las mismas. De todos modos puede necesitarse extender este modelo; por ejemplo se deben incluir los elementos necesarios para especificar la clase **`InformedeAuditoría`**, ya que las actividades de auditoría incluyen la gestión de informes para cumplir la legislación vigente en cuanto a protección de datos. También es necesario definir las equivalencias entre los conceptos de auditoría manejados por HISA y los manejados por otros estándares, como el EN13606, en los que pueden estar basados algunos de los sistemas a integrar.

6.4.8. Modelo de comunicaciones

En la extensión propuesta en este trabajo pertenecen al modelo de comunicaciones no sólo las tareas de interconexión de la arquitectura con elementos externos a la misma, sino también las actividades que puedan ser necesarias para la correcta comunicación de elementos dentro de la arquitectura. En este sentido, la definición de protocolos de comunicación entre componentes o de distintos mecanismos de acceso a la arquitectura pertenecería a este conjunto de tareas, así como la gestión de condiciones para la vinculación de interfaces, como por ejemplo de calidad de servicio.

En cuanto a la comunicación con elementos ajenos a la arquitectura, hay que considerar que la forma de comunicación con los mismos no siempre tiene que estar basada en mensajes, como en principio la norma señala, sino que pueden existir otros procedimientos. En cualquier caso existirán componentes que actúen de pasarela o puente entre la arquitectura y los elementos externos. Para la comunicación basada en mensajes, un buen referente para la definición de los componentes del modelo es el EN13606-5, donde se definen los formatos de los mensajes para el intercambio de historias clínicas, aunque hay siempre que considerar que en una arquitectura sanitaria no toda la información será referida a la historia clínica y por tanto hay que extender este modelo.

Para las comunicaciones dentro de la arquitectura se necesita un modelo, independiente a la tecnología subyacente, que ayude a representar los protocolos de comunicación, o flujos de trabajo, entre componentes. Siguiendo la terminología ODP, este modelo debe considerar que la interfaz de cada componente se define con una firma, un comportamiento y un contrato de entorno. La ontología utilizada para especificar estos términos (la representación de calidad de servicio requerida para realizar vinculaciones, los procesos del componente, etc) podría considerarse parte del modelo de información, aunque evidentemente la especificación concreta de cada componente forma parte del modelo computacional y es por ello que se presentará precisamente dentro de la vista computacional.

En lo referido a los mecanismos de acceso a la arquitectura hay que hacer consideraciones como el tipo de terminal utilizado, la red de acceso, la calidad de servicio requerida, etc. Factores que pueden influir mucho en la elección de los componentes a utilizar y el comportamiento de los mismos cuando hay que ofrecer

determinada aplicación al usuario. El modelo de información debe proveer de mecanismos para definir todos estos factores.

6.4.9. Conclusiones

El cambio más profundo que se propone en este trabajo es el de paradigma de modelado de la vista de información. En lugar de considerar la normalización de un modelo de información estático, se propone la definición de una ontología que permita cambiar o extender el modelo a medida que los conceptos o la cobertura del dominio lo hagan. En este sentido, aunque se ha representando el modelo de información con diagramas de clases, para utilizar una representación visual fácilmente comprensible, la ontología debe describirse con algún lenguaje de ontologías.

En cuanto a los modelos que cubren cada grupo de actividades, se ha ampliado el modelo para actividades semánticas, incluyendo conceptos de ontologías, anotaciones y reglas, y se señala que este modelo seguirá creciendo. El modelo para actividades de sujeto de cuidado se amplía más adelante en esta Tesis Doctoral y por eso sólo se señalan algunas consideraciones generales. En cuanto a los demás modelos, se proponen ampliaciones o modificaciones en base a otras normas más centradas en temas concretos, como el EN13606 o los servicios de CORBAmed, y en la consideración de un modelo federado.

Todos los modelos u ontologías están ligados a través de relaciones y todos son tratados como ontologías “vivas”, en el sentido de que se pueden añadir y extender clases y relaciones, se pueden fusionar ontologías, establecer equivalencias de conceptos de unas a otras, etc. Este mecanismo de definición de la ontología común facilita la incorporación a la federación de sistemas con distintos modelos de información ya que, como se indicó al principio, la premisa considerada es que siempre existirán sistemas heterogéneos y la finalidad del estándar es facilitar la integración.

6.5. Extensión de la vista computacional

6.5.1. Introducción

Según HISA, la especificación del modelo computacional debe ser independiente de las tecnologías, herramientas o soluciones adoptadas para la implementación de la arquitectura sanitaria. Al mismo tiempo debe ser suficientemente formal, completa y no ambigua para permitir a los implementadores un diseño eficiente del sistema y su

realización en una tecnología concreta. Esto es totalmente conforme al estándar ODP, en el que se destaca que el punto de vista computacional está directamente relacionado con la distribución del procesamiento, pero no con los mecanismos de interacción que hacen posible que se produzca la distribución.

La especificación computacional descompone el sistema en componentes que realizan funciones individuales e interactúan en interfaces bien definidas. Proporciona así la base para decisiones sobre la manera de distribuir los trabajos que habrán de realizarse. Esta distribución de procesado puede realizarse porque los mecanismos de comunicación que la soportan estarán definidos en la especificación de ingeniería.

Una normalización exhaustiva de cada componente aseguraría la interoperatividad; sin embargo, debido a la amplia cobertura de HISA y a su filosofía de universalidad dentro del marco sanitario, normalizar cada uno de los componentes puede resultar inviable, sobre todo considerando que el número de éstos aumenta a medida que crecen los requisitos de la organización o se incorporan nuevos dominios. En cuanto a este último punto, una normalización estricta de las interfaces dificulta la integración de sistemas ya desplegados que tienen que adaptarse a los criterios de la arquitectura.

Por tanto, el propósito de la especificación de la vista computacional en HISA no debe ser la especificación completa y fija de todos los componentes e interfaces que pueden ser necesarios para cumplir cualquier requisito que surja en cualquier organización sanitaria. Sólo debe especificar un conjunto de características identificadas como fundamentales y comunes en todas las organizaciones sanitarias y que deben ser satisfechas en el modelo computacional implementado en la arquitectura.

La principal aportación de esta Tesis en cuanto al desarrollo del modelo computacional es facilitar el arduo trabajo de normalización. Aquí se considera que el mayor esfuerzo debe ser la normalización del lenguaje computacional que tiene que permitir la descripción de interfaces, componentes y las capacidades, o servicios, que éstos ofrecen para tener un entendimiento completo y no ambiguo de los mismos. De este modo se facilita la incorporación de nuevos componentes y/o capacidades a medida que se integren nuevos dominios o surjan nuevos requisitos en la organización.

Para mantener la conformidad con ODP esta descripción debe ser independiente de la tecnología subyacente. Por otro lado, dado que desde el primer momento HISA se piensa como una particularización de ODP para el dominio sanitario, es importante

mantener la misma terminología empleada en la norma de referencia. De este modo cada componente se especificará como una o varias interfaces que se describen en términos de firma, comportamiento y contrato de entorno. De manera que es necesario definir las interacciones de cada interfaz, pero también sus reglas de comunicación y las condiciones que es necesario cumplir para establecer el vínculo con su interfaz complementaria y llevar a cabo las interacciones descritas en la firma de interfaz.

En cuanto al término servicio es necesario aclarar que la función de intermediación descrita en ODP (X.950 [146]) se encarga de facilitar el ofrecimiento y el descubrimiento de instancias de interfaces que proporcionan capacidades o servicios de determinados tipos. Es decir, que se entiende por servicio una capacidad que determinado componente ofrece a través de una interfaz. Sin embargo, dado que las interfaces pueden ser de tipo señal, operación o tren, no se puede entender que el uso de un servicio implique la adopción del paradigma cliente/servidor en dicha interfaz. Anunciar una capacidad u ofrecer un servicio se denomina «exportación». Concordar necesidades o descubrir servicios se denomina «importación». La exportación y la importación facilitan el descubrimiento dinámico del servicio adecuado para cubrir una necesidad y la posterior vinculación de interfaces complementarias para el logro del objetivo perseguido.

Normalizar el lenguaje de especificación de componentes y capacidades y conseguir que éste sea suficientemente flexible aporta dos ventajas fundamentales:

- Permite expresar todas las características de las interfaces y sus requisitos de vinculación de forma comprensible a personas o agentes software de forma independiente a la tecnología subyacente.
- Proporciona un mecanismo para la clasificación de capacidades o servicios que facilita las tareas de exportación e importación, y el establecimiento de vínculos entre interfaces. Es decir, se facilita la automatización del descubrimiento, invocación y composición de servicios.

Considerando este punto de partida, los mayores esfuerzos de HISA en la normalización de la vista computacional deberían centrarse en:

- Definir un lenguaje de descripción de componentes y capacidades, o elegir alguno de los existentes extendiéndolo, en caso de ser necesario, para describir los conceptos de ODP no considerados en el mismo. Este lenguaje debe incluir

la formalización de los mecanismos de expresión del contrato de entorno y de la calidad de servicio.

- Formalizar una taxonomía de servicios dentro de una organización sanitaria que permita una eficiente exportación e importación de los mismos a través de objetos mediadores. La normalización de esta clasificación de servicios debe apoyarse en las capacidades básicas necesarias en una organización sanitaria, pero se debe facilitar la extensión de tipos de servicios. El uso en el lenguaje de descriptores de servicios específicos del dominio sanitario, que además puedan extenderse sin dificultad, facilita esta tarea.

La función de depositario de tipos de ODP (X.960 [147]) contempla los conceptos necesarios para realizar adecuadamente estas tareas de descripción, y por tanto el lenguaje seleccionado debería estar basado en la ontología descrita en esta norma o al menos facilitar las correspondencias con la misma.

En cuanto a la clasificación de componentes, puede estar orientada por las últimas versiones de HISA donde por un lado se distinguen distintos tipos de componentes:

- **Básicos o de bajo nivel:** Permiten la introducción, consulta o modificación de instancias de los conceptos descritos en el modelo de información subyacente. Así, para cada clase del modelo de información la arquitectura debe proveer de un conjunto de métodos que permitan realizar estas tareas.
- **Complejos o de alto nivel:** Implementan transacciones de negocio más complejas, más cercanas a las actividades de usuario. Estas prestaciones más avanzadas se consiguen gracias a la interacción con otros componentes de acuerdo a las reglas específicas de la organización.
- **De utilidad general:** Que implementan funciones que son útiles para la gestión de todo el sistema.

Por otro lado HISA propone una clasificación muy simple de componentes que se corresponde con la descomposición realizada en actividades. Sin embargo es necesario considerar que existirán componentes que se utilizarán en distintos grupos de actividades, en especial aquellos que sean componentes básicos.

En esta Tesis, para poner de manifiesto la importancia del modelo federado, además de clasificar los componentes como básicos y complejos también se clasificarán como

autónomos y federados. Mientras que los autónomos manejan únicamente información de un sistema, los federados pueden manejar información perteneciente a varios dominios, por lo que pueden tener que resolver problemas de reconciliación de instancias, cosa que no ocurre en los autónomos.

A continuación se revisan los distintos conjuntos de actividades y se proponen algunas ideas que deben ser consideradas en la normalización de la vista computacional.

6.5.2. Lenguaje computacional

La normalización de un mecanismo de descripción de componentes que refleje todas las posibilidades que ODP permite, considerando además las particularidades del entorno sanitario y del modelo federado que prima en estas organizaciones, facilitar la interoperatividad y la extensión de componentes sin la necesidad de una normalización rigurosa y cerrada de interfaces. El lenguaje de descripción seleccionado se puede apoyar en esfuerzos previos de estandarización de lenguajes de descripción de la vista computacional (como TINA-ODL [133]), en otros lenguajes de especificación de servicios o, como se propone en esta Tesis, se puede utilizar un lenguaje de ontología de propósito general para describir la ontología ODP. Esto facilita la especificación de cualquier componente utilizando herramientas de gestión de ontologías ya existentes, simplemente editando las instancias de los conceptos representados en la ontología ODP.

Actualmente los mayores esfuerzos de investigación en programación distribuida están muy orientados a las tecnologías de la Web Semántica. Incluso miembros de otras comunidades, como la GRID semántica o los agentes inteligentes, intentan converger con estas técnicas dado el gran número de investigadores involucrados en el desarrollo de la Web Semántica. Este hecho puede influir en la elección de alguno de los lenguajes creados para la especificación de servicios Web semánticos, como WSDL-S [148], OWL-S [94-96] o WSML [149], como lenguaje propuesto por HISA. Sin embargo conviene tener en cuenta ciertas consideraciones:

- Dado que HISA es independiente de la plataforma subyacente, no es necesario que se utilice servicios Web como tecnología de integración.
- Sea cual sea el lenguaje de especificación de componentes utilizado en la materialización de la norma debe permitirse que la ontología del dominio pueda

estar expresada en cualquier lenguaje de ontología, e incluso se pueden utilizar varios.

- La comunicación entre agentes no tiene por qué estar regida por el paradigma cliente-servidor, y por tanto las interacciones no siempre son del tipo operación, como ocurre en los servicios Web.
- El lenguaje utilizado puede requerir extensiones para la especificación de algunas características de ODP no consideradas en los servicios Web.
- Es necesario normalizar una clasificación de componentes y capacidades de la arquitectura sanitaria para facilitar las tareas de descubrimiento de los mismos.
- Es necesario normalizar los mecanismos de descripción del contrato de entorno, que no están normalizados en los lenguajes de descripción de servicios. Aunque sí se consideran los términos precondición y efecto, o resultado, la forma de expresarlos no está normalizada y queda fuera del objetivo los mismos. Sin embargo existen otros estándares como SWRL [93] que están pensados para la descripción de reglas y que podrían utilizarse como punto de partida. En relación a este punto es necesario considerar que las reglas que rigen el uso de componentes pueden basarse en la gestión de estados y que la tecnología de servicios Web no maneja estados globales.

En esta Tesis Doctoral se propone que la tarea de normalización del lenguaje computacional sea abordada en dos pasos: la definición de la ontología ODP y la incorporación de los términos específicos del dominio sanitario.

Ontología ODP

Primero será necesaria la representación de la ontología de descripción de componentes y capacidades o servicios de ODP utilizando un lenguaje de propósito general para la especificación de ontologías, por ejemplo OWL. Esto sería el equivalente a lo que OWL-S es a los servicios Web semánticos, pero con la salvedad de que esta ontología contiene todos los conceptos descritos en la función de depositario de tipos (X.960 [128]), que resume los conceptos más importantes de ODP, y en la función de intermediación (X.950 [126]), necesaria para la publicación y descubrimiento de capacidades. Esta ontología está publicada en la dirección http://trajano.us.es/~isabel/EHR/ODP.owl.

Para facilitar la representación de la misma se han obtenido varios diagramas que muestran las relaciones más importantes reflejadas en la ontología, equivalentes a los incluidos en la norma X.960. Estos diagramas se han generado automáticamente a partir de la ontología expresada en owl utilizando "Ontoviz" [150] ejecutado sobre Protègè [151] y el software de libre distribución "Graphviz" [152].

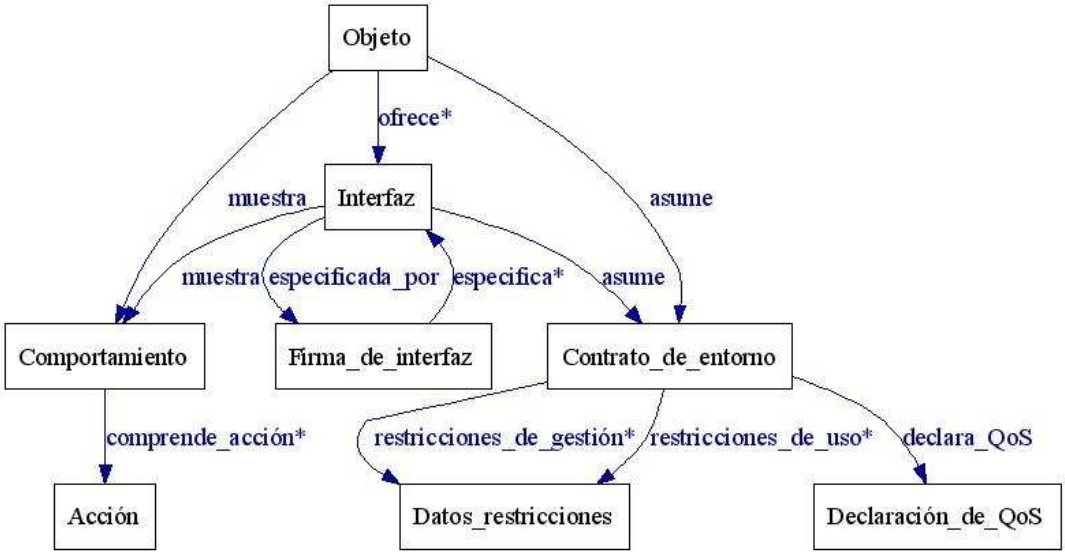


Figura 37. Clases objeto computacional e interfaz

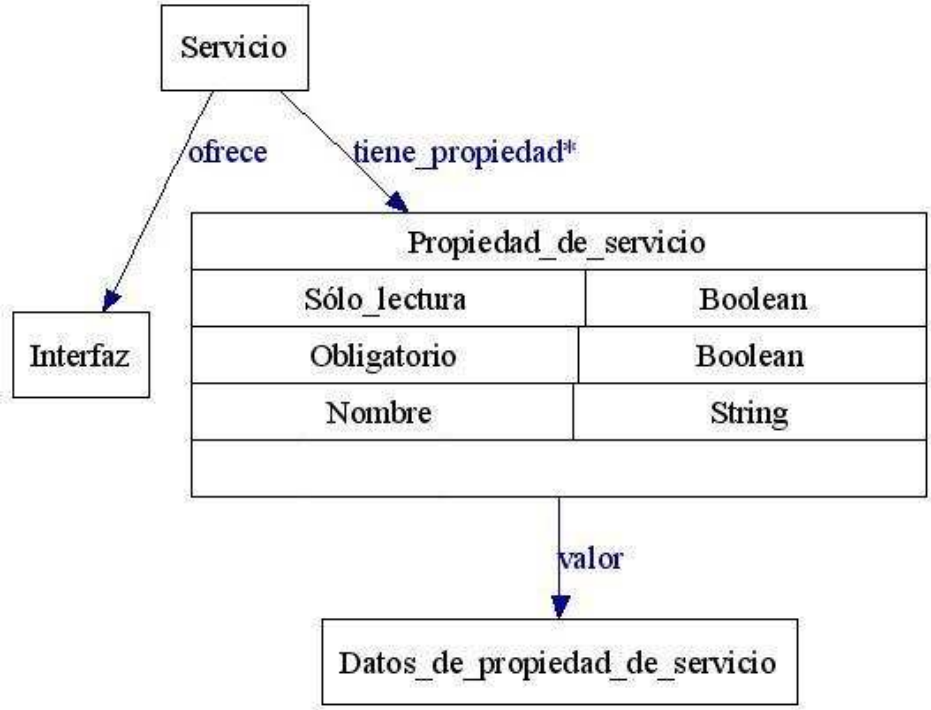


Figura 38. Clase servicio

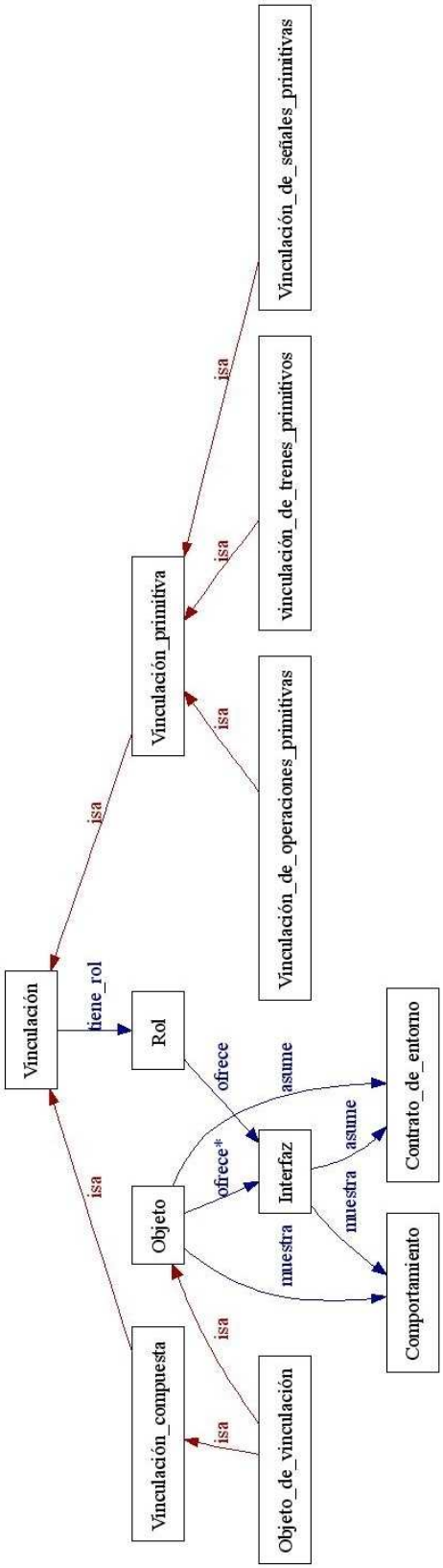


Figura 39. Subtipos de vinculación

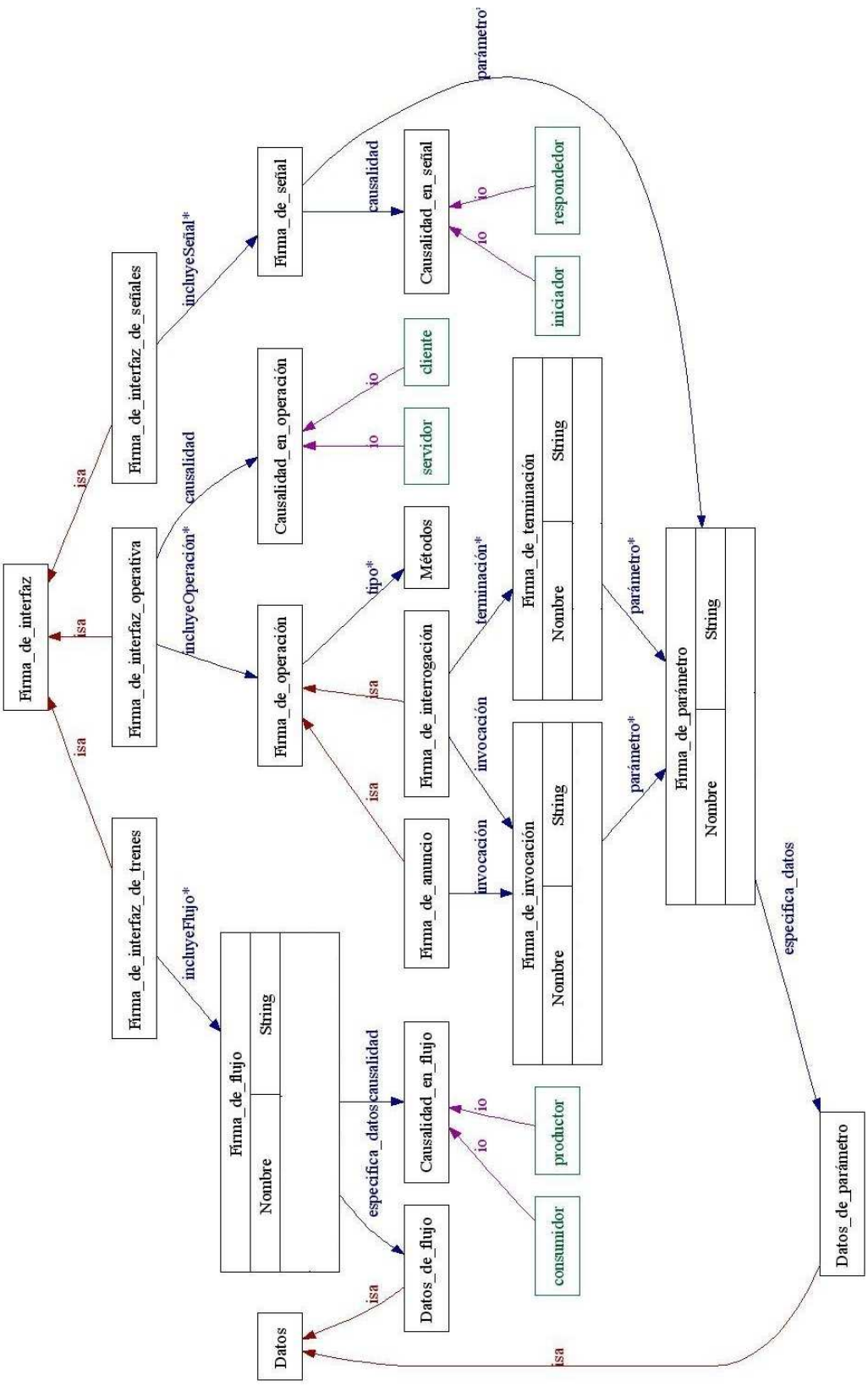


Figura 40. Clases de firma y datos

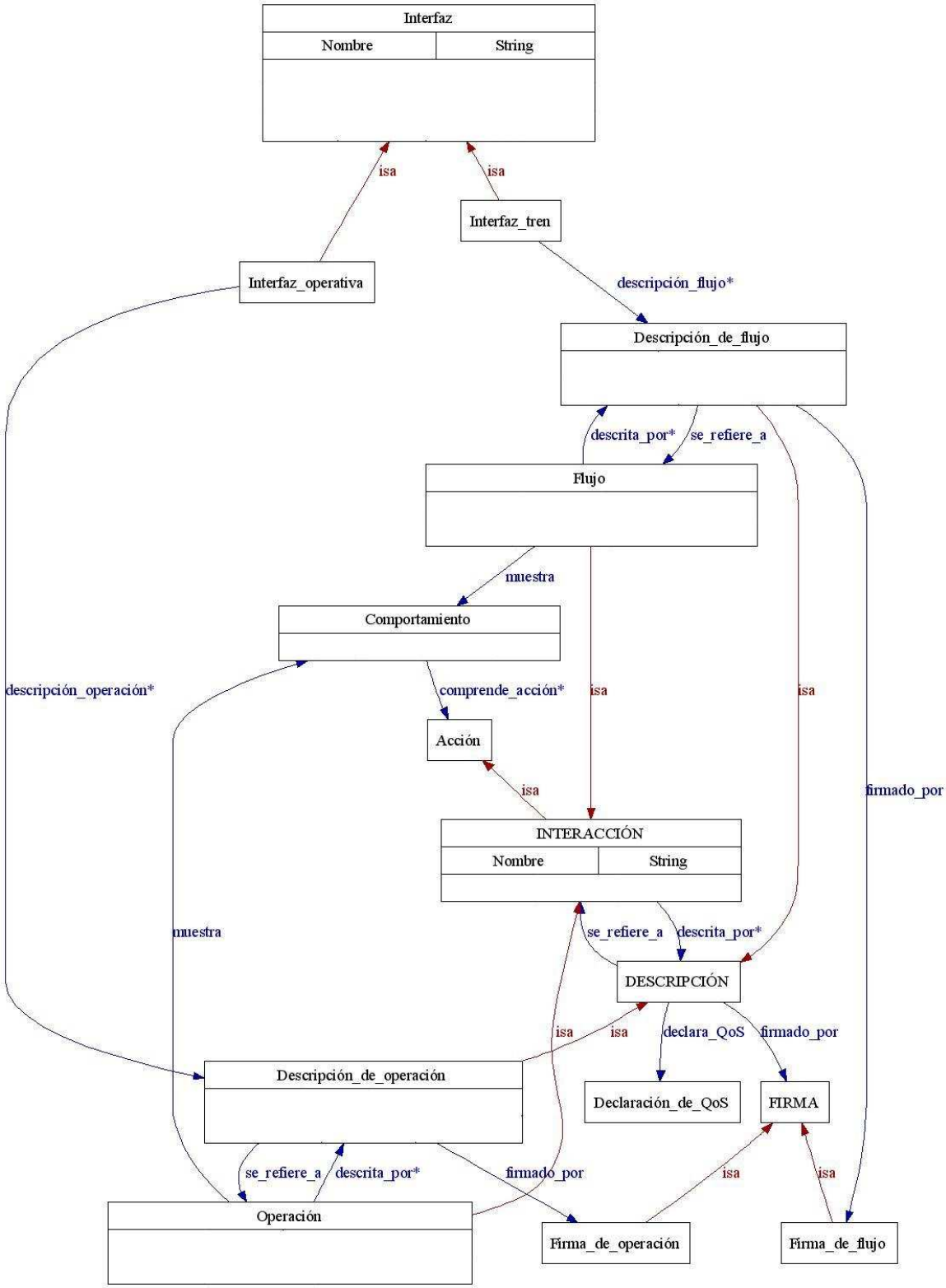


Figura 41. Tipos de interfaz para la vinculación

Algunas de las ventajas que este mecanismo aporta son:

- Al estar especificado el lenguaje de descripción como una ontología se facilita la extensión del mismo, y por tanto la normalización en HISA consistirá en una extensión que cubra el dominio sanitario.
- Las extensiones podrían ser realizadas también en implementaciones concretas.
- Se facilita el establecimiento de correspondencias entre descripciones de componentes basadas en otras ontologías, como OWL-S.
- Se pueden utilizar herramientas ya existentes para la edición de descripciones de componentes.

En el contexto tecnológico de la Web Semántica se están realizando grandes esfuerzos en la actualidad y muchos de los avances podrían ser utilizados en el dominio sanitario, independientemente de la tecnología utilizada para la implementación.

Extensión de la ontología de componentes con los términos del dominio sanitario

El siguiente paso es la extensión de esta ontología con los conceptos necesarios en el dominio sanitario. En particular sería necesario:

- Normalizar los atributos de descripción de capacidades o servicios para que cubran las necesidades del entorno sanitario. Dado que en HISA los servicios están clasificados según el grupo de actividades al que pertenecen, los descriptores deben permitir la consideración de esta clasificación. Estos atributos facilitarán la selección de componentes.
- Incluir en la ontología los métodos básicos propuestos en la actual versión de HISA, y posiblemente extenderlos, como subclases de los conceptos apropiados. En concreto HISA los modela todos como operaciones, aunque esto no tendría por qué ser así.
- Incluir las interfaces básicas o complejas imprescindibles en el dominio sanitario, así como las capacidades que deben soportar. Es necesario incluir las que ya están descritos en la versión actual de HISA pero también facilitar la extensión.
- Incluir en la ontología los servicios de propósito general propuestos en la versión actual de HISA.

Las figuras que se presentan a continuación representan algunos de estos conceptos que han sido ya incluidos en la ontología. La extensión es sumamente fácil utilizando algún editor de OWL como Protégé [151].

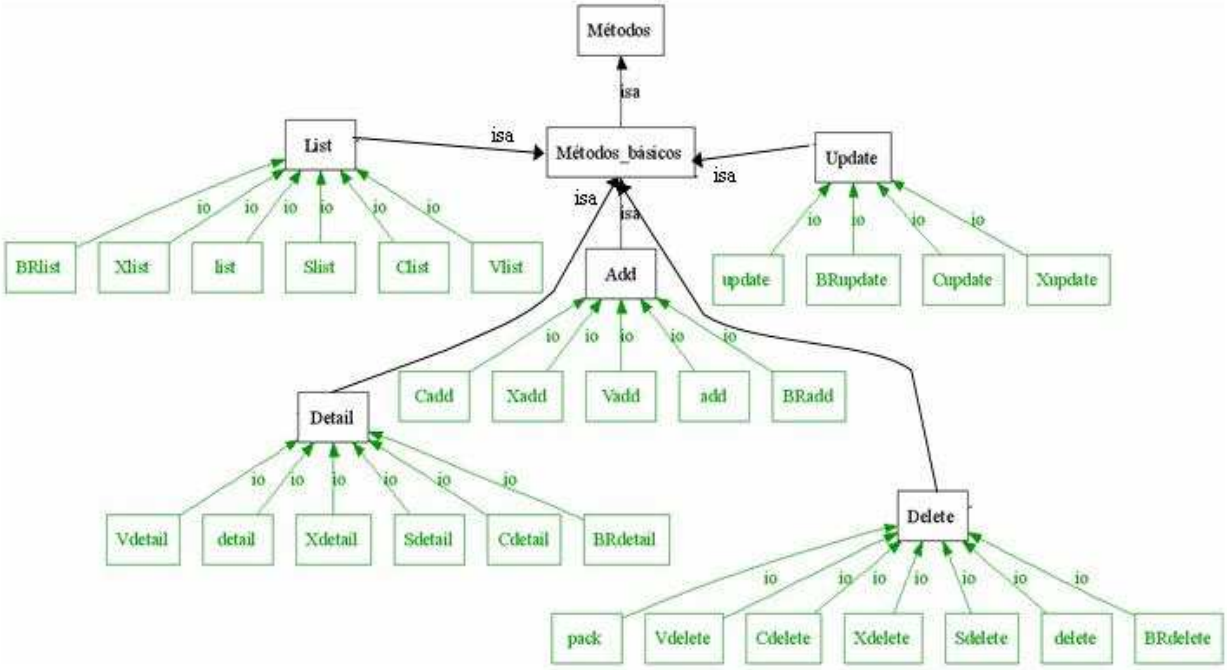


Figura 42. Datos para la clasificación de métodos básicos

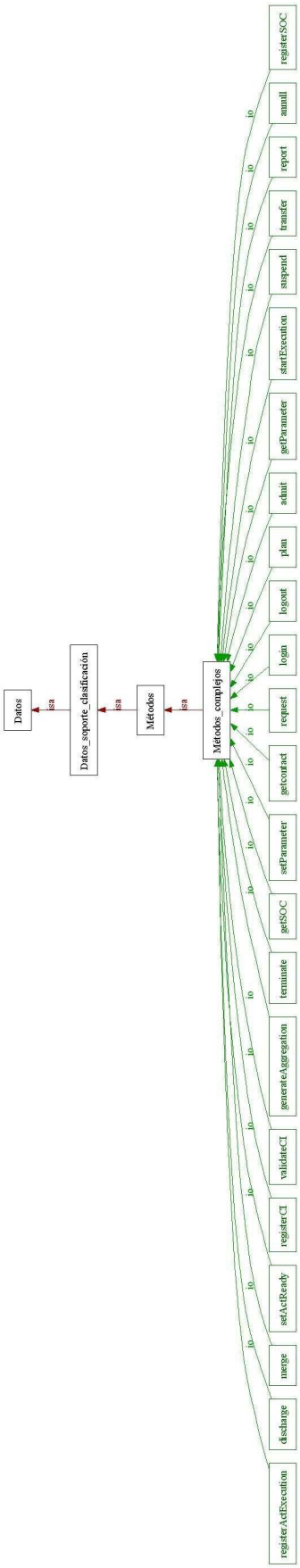


Figura 43. Datos para la clasificación de métodos complejos

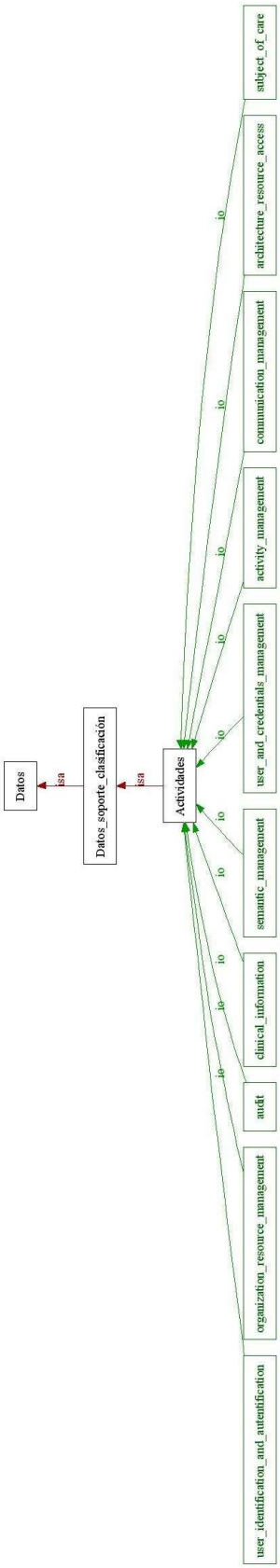


Figura 44. Datos para la clasificación de actividades

6.5.3. Componentes semánticos de la arquitectura base

Las funciones de estos componentes son de gestión de los demás, por lo que pueden considerarse parte de la arquitectura subyacente a HISA. Sin embargo es interesante presentarlos aquí ya que su existencia condiciona la forma de especificar el resto de componentes que sí se consideran parte de la arquitectura de sistemas sanitarios.

Se describen estos componentes de forma independiente a la tecnología concreta que se utilice, simplemente como una serie de funciones que si existen en una plataforma de esta naturaleza permiten una gestión semántica de componentes muy flexible, facilitando el descubrimiento, invocación y composición de los mismos.

Si HISA se soporta sobre una plataforma con gestión semántica, existirán componentes o agentes que utilizando la descripción de los componentes de la arquitectura sanitaria puedan realizar su gestión semántica. Éstos sirven de apoyo a la interoperatividad funcional y semántica. Se consideran parte de la plataforma subyacente porque se centran en conceptos genéricos de arquitectura y no en los del dominio sanitario propiamente dicho, por ello no se consideran parte de ningún conjunto de actividades y ni siquiera tendrían por qué ser descritos en HISA. En este trabajo se introducen para mostrar las ventajas que el uso de este tipo de tecnologías podría aportar a una arquitectura de sistemas sanitarios. Las funcionalidades más básicas son:

- **Componentes para la publicación de componentes:** sería el concepto equivalente a los objetos de intermediación descritos en X.950 o, en una tecnología concreta como los servicios Web, el UDDI. Permiten gestionar un repositorio de descripciones de componentes incluyendo tareas de registro y consulta de los mismos. Esta información incluirá la descripción de las capacidades del componente y también información funcional del componente que se necesite para el establecimiento de vínculos entre interfaces y la interacción con el mismo. El lenguaje de descripción utilizado en cada implementación dependerá de la tecnología elegida, pero los componentes y servicios en HISA deben describirse de forma suficientemente abierta para que la traducción a cualquier lenguaje de especificación de componentes pueda realizarse de forma sencilla. El lenguaje descrito en este trabajo ha sido desarrollado con esta filosofía.

- **Componentes para el registro de componentes:** interaccionan con los de publicación para registrar capacidades en el repositorio, sería la tarea equivalente a la exportación descrita en la X.950. Generalmente se podría contemplar la posibilidad del autoregistro, es decir, que todos los componentes incluyan esta interfaz para registrarse en los componentes de publicación, pero no siempre tiene que ser así y pueden existir componentes específicos para realizar esta tarea.
- **Componentes para la selección de componentes:** interaccionan con los de publicación para consultar el repositorio y seleccionar los componentes adecuados al objetivo marcado; sería la tarea equivalente a la importación descrita en la X.950. Una vez elegido el servicio y el componente que lo publica, la firma de interfaz puede guiar las interacciones que ocurran entre las interfaces vinculadas. Para objetivos complejos, no alcanzables con la interacción con un único componente, pueden existir componentes vinculadores que interaccionan con varios agentes para ofrecer una interfaz más compleja. Éstos pueden apoyarse en el uso de **componentes asistentes a la composición**, que ayudan a la creación de objetos vinculadores que a partir de la interacción con otros agentes ofrecen nuevas prestaciones para cumplir el objetivo marcado. La automatización de estas tareas, además de la descripción exhaustiva de las interfaces de componentes necesita una descripción muy precisa de necesidades, contratos de entorno y requisitos de QoS [153,154].

En definitiva, aunque estos componentes no se consideren parte de HISA sino de la arquitectura subyacente, sí que permiten identificar qué tipo de información semántica debe incluirse en la definición de componentes. El lenguaje de descripción sí debe estar especificado en HISA, como se dijo anteriormente, pero deja de ser necesaria una normalización explícita y rigurosa de cada una de las interfaces de estos componentes. De este modo será sencillo extender la arquitectura con nuevas funciones y se facilitará la integración de sistemas siguiendo el modelo federado.

Adoptando este paradigma, el mayor esfuerzo en la vista computacional de HISA debería ser la normalización del lenguaje de descripción de componentes y la identificación de los componentes y servicios básicos que la arquitectura sanitaria debe incorporar. Sin embargo, puede resultar interesante dar la descripción exhaustiva de algunos componentes, o apuntar hacia interfaces ya normalizadas, para poder guiar a los

desarrolladores de nuevos sistemas en la forma de implementar los componentes a integrar en la arquitectura.

6.5.4. Componentes de gestión semántica

Al igual que con la vista de información, comenzamos analizando las modificaciones y extensiones propuestas en los componentes de gestión semántica, ya que son fundamentales para las tareas de integración en un modelo federado y constituyen una importante aportación de este trabajo de investigación.

La tarea fundamental de estos componentes es manejar las instancias de las clases relacionadas con la gestión semántica ya presentadas en la vista de información, y por tanto tratan específicamente conceptos de la ontología del dominio de la arquitectura sanitaria. En definitiva, podrían considerarse los objetos que permiten la gestión dinámica del modelo de información. Este concepto es similar a la función de depositario de tipos especificada en la X.960, parte también de la norma ODP. Principalmente se podrán encontrar:

- **Componentes gestores de ontologías:** entre ellos estarán los que permitan modificar la ontología del dominio, es decir el modelo de información, facilitando la introducción, borrado y modificación de clases, propiedades y restricciones. También existirán componentes que simplemente permitan consultas a ontologías, para conocer el modelo de información compartido en un determinado instante. Van a ser reutilizados por agentes especializados en el establecimiento de las relaciones entre términos de distintas ontologías, que serán de gran apoyo a la integración. Básicamente son componentes autónomos que permiten la automatización, o semiautomatización, del establecimiento de correspondencias entre conceptos locales de un sistema y los federados, encontrando las equivalencias y diferencias entre clases y propiedades de ambas. Generalmente serán utilizados por componentes autónomos para que éstos puedan relacionarse con otros en términos de la ontología federada. La normalización de las interfaces de estos componentes pueden basarse en sistemas actuales de gestión de ontologías como JENA [155] o Protègè [156], que permiten tanto la modificación como la consulta de ontologías e incluso tienen funciones de fusión de ontologías. De hecho podrían reutilizarse estos componentes en la arquitectura ya que son de libre distribución. Dado el carácter

genérico con el que se plantean estos componentes también podría considerarse que no son parte de HISA sino de la arquitectura de integración básica en la que HISA se apoye.

- **Componentes gestores de arquetipos:** se pueden considerar un subgrupo de los anteriores, más específicos del dominio sanitario, y que harán uso de los más generales. Ya que en el modelo de información se ha considerado la introducción del concepto de arquetipo tal y como se especifica en el openEHR y en el EN13606, es conveniente que la arquitectura cuente con agentes que los gestionen adecuadamente. Los componentes más básicos serían los que tengan interfaces para la creación, almacenamiento, gestión y/o consulta de arquetipos. Para facilitar la integración pueden existir componentes de traducción de arquetipos, que permitan traducir las restricciones impuestas en los conceptos de la ontología federada a restricciones para las ontologías locales. También pueden existir componentes que descubran el arquetipo al que se corresponde cierta instancia de información y cualquier otra gestión de arquetipos que sea necesario realizar.
- **Componentes gestores de terminologías y tesauros:** son los únicos considerados actualmente en HISA. La interfaz detallada puede estar basada en alguna especificación ya existente, como el servicio de consulta de léxico (Lexicom query service, LQS) de CORBAmed. Básicamente tendrán interfaces de consulta a sistemas de términos y códigos que, o bien devuelven el código correspondiente a algún concepto en determinado sistema de términos, o bien realizan el proceso contrario. Pueden existir tareas más complejas, como la traducción de un sistema a otro, la consulta de términos en un idioma específico, etc.
- **Componentes de gestión de reglas y razonadores:** la funcionalidad básica es soportar la creación y gestión de reglas y realizar las inferencias oportunas en base a los hechos acontecidos en el sistema. Servirán para disparar acciones en distintos componentes o para inferir nuevo conocimiento a partir del ya considerado en el sistema. En este sentido serán el soporte de componentes más complejos como los de apoyo a la decisión médica.

- **Componentes de gestión de anotaciones:** las tareas de estos componentes se centran en facilitar la creación y almacenamiento de anotaciones y en manejar los vínculos con la información a la que complementan. Además se encargan de la recuperación, ayuda a la interpretación, presentación y gestión de las anotaciones.

Las anotaciones matizan la semántica de la información a la que están ligadas, la hacen más específica o incluso pueden llegar a cambiar el significado de la misma. A partir de las anotaciones se puede llegar a decidir qué componentes están más capacitados para gestionar adecuadamente la información que cualifican o cuáles pueden proporcionar un conocimiento más amplio o específico a partir de las anotaciones adjuntas.

Cuando las anotaciones se refieren a observaciones clínicas, se establecerán relaciones entre estos componentes y los de gestión de información clínica. En este caso las anotaciones suelen aportar conocimiento sobre la parte del cuerpo a la que afecta determinada observación, la relevancia de la misma en ciertas actividades, la actividad en la que se generó o los códigos correspondientes en distintos sistemas de clasificación.

La gestión semántica de términos combinados puede considerarse parte de las tareas de los componentes de anotación. Podrían existir componentes especializados en manejar combinaciones del tipo calificación, modificación, negación o especificación que podrían colaborar con los agentes gestores de terminologías y tesauros para entender el significado preciso de los términos anotados y gestionarlos adecuadamente.

Teniendo en cuenta que la mayoría de la información clínica se maneja en formato texto, los componentes que gestionan texto anotado requerirán una mención especial.

6.5.5. Componentes relacionados con el sujeto de cuidado

Se pueden distinguir dos tipos de componentes:

- **Componentes del sujeto de cuidado:** se encargan de la gestión del sujeto de cuidado, usualmente pacientes. Un buen referente para desarrollar estos componentes es el servicio de identificación de personas [29] de CORBAmed. El siguiente capítulo de este trabajo de tesis se centra precisamente en estos componentes. Algunas de las acciones básicas a realizar son: registro de una persona, consulta de datos personales, modificación o introducción de nuevos

datos, fusión de dos personas que resultan ser la misma, etc. Todo ello considerando el modelo federado, es decir que el registro puede realizarse localmente, que el paciente puede tener distintos identificadores en distintos dominios o que los valores de los atributos del mismo paciente en distintos sistemas pueden no coincidir.

- **Componentes para la gestión de contactos:** En lo relativo a la gestión de contactos sanitarios, las normas CEN que se están usando como referente principal no entran en estas consideraciones, sin embargo sí lo hacen algunos otros estándares bien aceptados como DICOM. La norma define los estados en los que se puede encontrar un contacto y las posibles transiciones entre éstos, como se muestra en la figura 45.

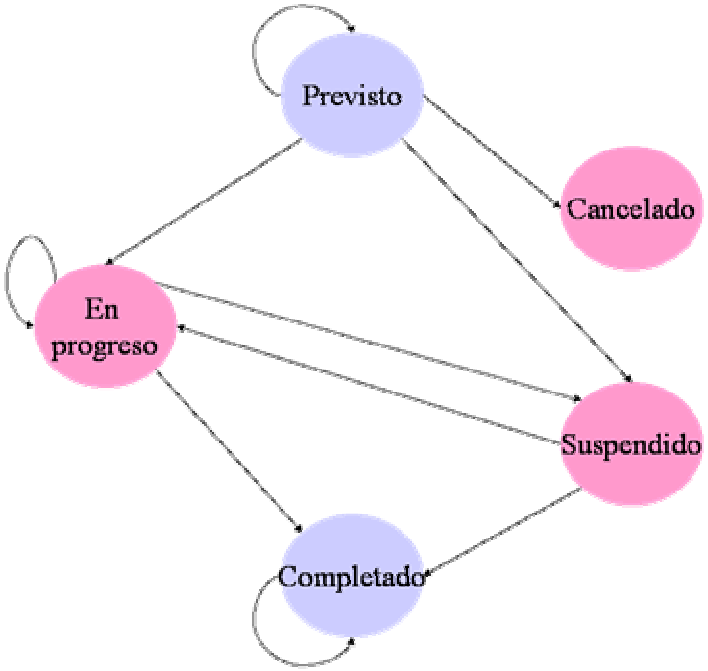


Figura 45. Estados posibles de un contacto

Los componentes tienen que facilitar la gestión de un contacto al menos en los siguientes términos:

- Creación o inclusión en la agenda
- Identificación y recuperación
- Modificación y consulta de su estado

Los componentes de gestión de contactos pueden interaccionar con componentes de otros grupos de actividades. Por ejemplo un contacto pasa de **`previsto`** a **`en progreso`** en el instante en que los recursos necesarios para llevar a cabo ese contacto están disponibles. Durante todo el estado **`en progreso`** el recurso está ocupado, pero en el momento en el que se pasa a **`completado`** el recurso vuelve a estar disponible. Habitualmente los contactos forman parte de una actividad, por lo que también existirán interacciones con los componentes de gestión de actividades.

Los componentes que soportan estas actividades van a ser reutilizados para ofrecer otras prestaciones más complejas, como veremos más adelante. A su vez ellos reutilizan componentes propios de otros grupos de actividades, en especial los de gestión semántica y seguridad, actividades muy transversales como se especificó anteriormente.

6.5.6. Componentes de información clínica

Como se vio en el modelo de información, el objeto fundamental en las actividades de gestión de información clínica es la observación clínica. De manera que los componentes que soporten este grupo de actividades deben proporcionar prestaciones para gestionar estas observaciones, desde su creación hasta su presentación al usuario. Es evidente que existirán numerosas interacciones entre estos componentes y los que pertenecen a otros grupos de tareas. A modo de ejemplo, para recuperar una observación puede necesitarse la identificación del paciente, una observación puede estar anotada, el resultado de un contacto es una observación y, por supuesto, las políticas de acceso a cada observación deben ser definidas y cumplidas.

Dado que la naturaleza de las observaciones es muy diversa, las tareas de gestión semántica serán especialmente utilizadas por los componentes de este grupo de actividades. Siguiendo el modelo dual y utilizando el concepto de arquetipo, una observación puede estar asociada a un arquetipo en el que se exprese la semántica concreta del mismo. Este arquetipo define la estructura de la información y las posibles restricciones en los valores. Así, los componentes encargados de tareas relativas a la información clínica interaccionarán con los de gestión de arquetipos. Es también importante poder conocer en qué actividad se generó la observación, la influencia de la misma en ciertas actividades, referencias a sistemas de clasificación o partes del cuerpo implicadas, etc, esta información puede adjuntarse a la clínica en forma de anotaciones,

de manera que también existirán vínculos entre los componentes de gestión de anotaciones y los de información clínica.

Podrían distinguirse tres actividades básicas a realizar sobre una observación y sus componentes asociados:

- **Componentes de adquisición y creación:** los componentes que cubren la adquisición y creación son muy diversos y muy próximos al usuario. Se pueden incluir en esta categoría diversos sensores biométricos, complejos sistemas de adquisición de imágenes o sencillos procesadores de texto. Todos estos componentes tienen que incluir interfaces para interaccionar con los componentes de gestión. Durante esta fase de adquisición y creación se pueden adjuntar anotaciones a la información clínica, para matizar su semántica o porque puedan ser de utilidad en la futura gestión de la misma, por ejemplo se puede anotar en qué actividad se genera, qué dispositivo la recoge, un código conforme a algún sistema de clasificación, la hora de la medición o las condiciones del paciente. Por otro lado, las observaciones pueden ser creadas conforme a arquetipos, con lo que la semántica de las mismas queda perfectamente recogida.
- **Componentes de gestión:** son los componentes básicos de este grupo de actividades y se podrían especializar en el almacenamiento, recuperación y modificación de la información clínica. El servicio que se puede utilizar como referente es el COAS [54] de CORBAmed, aunque existen otros que también se deben considerar y que están más centrados en tipos concretos de observaciones, como por ejemplo el modelo de servicios de DICOM [51] o el CIAS [55] de CORBAmed, también pensado para la gestión de imágenes. La ventaja de COAS es la propuesta de un servicio independiente del tipo de observación manejado. Siguiendo este modelo, en HISA se pueden definir los componentes básicos genéricos, para almacenar, consultar o modificar cualquier observación, y otros más específicos que apoyándose en los anteriores gestionen imágenes, pruebas de laboratorio o informes clínicos. Para determinar el componente adecuado a la gestión de una observación será necesario conocer el arquetipo en el que está basada y las anotaciones enlazadas a la misma; en este sentido las relaciones de los componentes de gestión de información clínica con los

componentes de gestión de arquetipos y de anotaciones deben ser descritas en detalle.

En HISA se considera que las observaciones tienen tres estados posibles; **`preliminar`**, **`confirmada`** y **`anulada`**. Los componentes de gestión permitirán conocer el estado de una observación y modificarlo cuando sea necesario.

En el modelo federado que se está presentando hay que considerar que, debido a que una observación compuesta está formada a partir de observaciones más simples, en los componentes federados que gestionen observaciones puede darse la circunstancia de que las observaciones que forman parte de una compuesta se obtengan desde distintos dominios. Sin embargo la interacción para gestionar observaciones con componentes autónomos o federados no debe diferir por este hecho. En cuanto a los posibles conflictos de entidad, la información clínica no debería duplicarse en el sistema; sin embargo esta condición no siempre se cumple, sobre todo cuando se intercambian informes entre sistemas de la arquitectura, ya que éstos suelen almacenarse localmente. Los componentes que gestionan información clínica se apoyarán en los que realizan tareas de control de versiones, gestión de responsabilidad sobre la información y auditoría para resolver estos conflictos.

- **Componentes de uso y presentación al usuario:** el componente adecuado para la presentación al usuario de una observación va a depender del contexto, que contemplará parámetros como el tipo de observación o el arquetipo en el que está basado, las anotaciones correspondientes, el dispositivo que se esté utilizando para la visualización, el uso concreto incluyendo la actividad, etc. No siempre la recuperación de una observación clínica tiene la finalidad de presentarla directamente al usuario, sino que éstas pueden utilizarse por componentes que realicen otras operaciones, como por ejemplo análisis estadísticos o epidemiológicos, elaboración de informes de cualquier naturaleza o ayuda a la decisión clínica. Por supuesto en cualquier caso todos estos agentes deberán incluir interfaces para interaccionar con los componentes de gestión de información clínica.

6.5.7. Componentes de gestión de actividad

En una organización sanitaria distintos tipos de actores llevan a cabo actividades durante las cuales se utilizan determinados recursos y se generan resultados. Estas actividades están relacionadas, directa o indirectamente, con las necesidades de los pacientes o con los requisitos generales de la organización. De manera que según el tipo de actividad, los resultados pueden ser información clínica sobre el paciente o información de interés en otros sectores de la organización. El ciclo de vida de una actividad puede ser modelado por el flujo de trabajo entre los actores, incluyendo los eventos y estados por los que pasa la actividad. Cada recurso utilizado durante el ciclo de vida implica un coste que depende del tipo de recurso y de la actividad.

Los conceptos relativos a la gestión de actividades son fundamentales en una organización sanitaria: la integración, monitorización y optimización del flujo de trabajo de los actores implicados en las actividades de la organización es imprescindible para mejorar la calidad de los servicios ofrecidos al cliente, la organización interna o la gestión de recursos.

Cuando una actividad es ejecutada por distintos agentes debe establecerse un protocolo de comunicación entre ellos. Generalmente este protocolo comienza con el envío de un mensaje de petición donde se indica en detalle los servicios solicitados y cualquier información de interés para llevar a cabo la actividad, como la urgencia en la ejecución o la identificación del que hace la petición. En actividades más complejas pueden verse implicados más de dos agentes, y las reglas de comunicación pueden cambiar según el tipo de actividad. En un determinado instante del ciclo de vida el sistema encargado de la actividad en ese momento debe identificar quién es el responsable y quién tiene permiso para realizar ciertas acciones e incluso pasar la actividad a otro estado.

Los componentes de gestión de actividad podrían clasificarse en:

- **Gestión de estados:** permiten conocer y cambiar en cualquier momento el estado de las actividades. Es fundamental la identificación inequívoca de la actividad para poder referirse a ella, tarea especialmente compleja si hay varios sistemas implicados en la misma actividad. En un instante determinado cada componente de este grupo debe saber quién está autorizado a realizar consultas o cambios de estado sobre una actividad determinada.

- **Gestión de la comunicación:** se encargan de enviar o recibir los mensajes del protocolo de comunicación entre los componentes implicados en la gestión de la actividad. Están muy relacionados con los de gestión de estados ya que generalmente un cambio de estado puede provocar el envío de un mensaje y viceversa.
- **Gestión de catálogos de actividades:** permiten conocer los tipos de actividades disponibles y dan información sobre la funcionalidad, recursos utilizados, protocolos de comunicación, flujo de trabajo, etc, de cada una de ellas. La consulta a estos componentes permite conocer los componentes implicados en el ciclo de vida de una actividad en particular. Podrían considerarse parte de éstos los componentes de Guías clínicas, que proporcionan información sobre la adecuada planificación del modo de actuar ante determinada patología y circunstancias. En general podría distinguirse entre la gestión de repositorios de protocolos clínicos, operacionales o administrativos; todos ellos guían a los demás componentes durante el ciclo de vida de una actividad.

En el caso de actividades de carácter asistencial, la mayoría de ellas están directa o indirectamente relacionadas con un contacto del paciente con la organización sanitaria. Esto indica que existirán vínculos de comunicación entre los componentes del sujeto de cuidado dedicados a la gestión de contactos y los dedicados a la gestión de actividades.

La finalidad de toda actividad es generar unos resultados; cuando se generan observaciones se debe interaccionar con los componentes de información clínica que se ocuparán de la gestión de éstas. La identificación de la actividad que ha dado lugar a la creación de una observación clínica puede aportar una mejor contextualización de la misma. Esto supone aumentar el conocimiento clínico que la observación aporta y por ello es conveniente que esta información se almacene en forma de anotaciones ligadas a la observación, que serán gestionadas por los componentes de anotación.

6.5.8. Componentes de recursos de organización

La gestión se refiere en este caso a los recursos de la organización sanitaria. Como se ha presentado en el punto anterior, durante una actividad generalmente se utilizan una serie de recursos. Esto muestra que estos componentes estarán estrechamente ligados a los de gestión de actividad, ya que en numerosas ocasiones el cambio de estado de una

actividad dependerá de la disponibilidad o no de los recursos necesarios para llevarla a cabo.

Por otro lado, dado que el gasto de la organización podría considerarse en función de los recursos consumidos o utilizados, la gestión de éstos está intrínsecamente unida a la gestión del gasto. De manera que la planificación de la organización puede basarse, en gran medida, en la consulta a estos componentes sobre la utilización realizada de los recursos disponibles.

Atendiendo a la sencilla clasificación de recursos presentada en la vista de información, algunos de los componentes que podemos encontrar dentro de este grupo podrían ser:

- **Componentes de gestión de recursos humanos:** permiten la gestión del personal ligado a la organización. Entre otras prestaciones pueden facilitar la localización de la persona con el perfil más adecuado para participar en determinada actividad. En este sentido pueden interactuar con los componentes de gestión de credenciales y con los de gestión de actividad. Serán reutilizados por los componentes administrativos encargados, por ejemplo, de la gestión de nóminas.
- **Componentes de gestión de material:** realizan sobre todo el control del material fungible. Indicarán cuándo es necesario hacer pedidos a los proveedores y elaborarán la lista de material a solicitar.
- **Componentes de control de agendas:** dado que la gestión de actividades está estrechamente ligada a la disposición de recursos, que es limitada, es necesaria una utilización óptima de los mismos y por ello la gestión de agendas, tanto para las tareas asignadas al personal como para las reservas realizadas sobre determinado recurso debe ser tarea de los componentes pertenecientes a este grupo. Éstos agentes interactuarán con los de gestión de actividades cuando sea necesario.

6.5.9. Componentes de seguridad y gestión de recursos

Componentes de usuarios, identificación y credenciales

Gestionarán los objetos referentes a la identificación de usuarios, autenticación y credenciales de los mismos. Serán el soporte de otros componentes como los de autorización de acceso a recursos de la arquitectura. Las interfaces de comunicación

pueden estar basadas en estándares previamente existentes en esta materia, como SAML (Security Assertion Markup Language) [157]. Al ser tareas muy comunes en distintos escenarios, si la plataforma que soporta HISA es adecuada pueden estar resueltas en la misma, de manera que la arquitectura sanitaria sólo debe particularizarlas para el dominio clínico. En este sentido, el mayor esfuerzo se realizará en la gestión de los modelos de información de manera que contemplen los datos necesarios en este escenario, mientras que las interfaces de comunicación con los componentes serán reutilizables casi en su totalidad.

Se podrían distinguir:

- **Componentes de gestión de usuarios:** se encargan de la creación, modificación y borrado de usuarios. Son locales en cada sistema, ya que en entornos federados se debe mantener la autonomía para que cada dominio gestione sus propios usuarios.
- **Componentes de autenticación:** a partir de algún mecanismo se verifica que el usuario es quien dice ser. Este mecanismo puede requerir la intervención del usuario, por ejemplo con la introducción de un identificador y una clave, o ser transparente al mismo, por ejemplo con el uso de certificados de clave pública.
- **Componentes de gestión de credenciales:** su tarea es proporcionar la información relativa al usuario que básicamente va a ser utilizada para tomar decisiones en cuanto a las autorizaciones del uso de recursos de la arquitectura. En entornos federados, las credenciales suelen ser locales a cada sistema, de manera que los componentes que las gestionan también lo son.
- **Componentes de gestión de identidades:** como se viene señalando, en entornos federados las tareas de gestión de usuarios y credenciales son complejas, ya que deben mantenerse locales a cada sistema y sin embargo cuando se necesita de la intervención del usuario debe facilitarse la transparencia al mismo. En particular, si el usuario tuviera que participar activamente en su identificación cada vez que acceda a uno de los sistemas de la federación, el uso de las aplicaciones facilitadas sería muy tedioso. Así, en un entorno federado son imprescindibles agentes gestores de identidades, de manera que el usuario se identifica sólo una vez y éstos se encargan de hacer la identificación en su nombre en los sistemas que sea necesario.

Es también importante destacar que aunque un usuario no siempre es una persona, ya que cualquier componente se puede considerar usuario de otro, ésta es la acepción más utilizada. En este caso hay que considerar que si se hace una buena descomposición del sistema, determinados componentes implicados en la gestión de personas, ya sean sujetos de cuidado u otros usuarios, pueden ser ampliamente reutilizados.

Componentes de recursos de arquitectura y autorizaciones

La tarea fundamental de estos componentes es decidir, en base a las credenciales de un usuario proporcionadas por los componentes anteriores, si tiene permiso o no para acceder a un recurso y realizar determinadas operaciones sobre el mismo. Servicios como el RAD [56] de CORBAmed pueden servir de base para desarrollar las interfaces de estos componentes. Del mismo modo que los anteriores, en determinadas plataformas pueden existir componentes que realicen estas tareas de seguridad y que podrán ser utilizados realizando las extensiones necesarias al dominio sanitario.

Componentes de auditoría

Los componentes básicos de este grupo de actividad serán los encargados de gestionar la información que se refiere a control de versiones, cambios y responsabilidades sobre la información almacenada en la arquitectura. Dado que el modelo considera estas estructuras de datos comunes a todos los objetos del sistema, van a ser componentes ampliamente reutilizados por todos los demás. También se pueden considerar aquí los componentes encargados de realizar los registros de actividad dentro de la arquitectura, conforme a la legislación vigente, así como los encargados de la elaboración de los informes de auditoría pertinentes. Estos componentes pueden ayudar a registrar o consultar información sobre la creación y modificación de observaciones clínicas u otros datos, los accesos, la creación de copias de respaldo, la gestión de soportes de entrada o salida o las incidencias de cualquier tipo.

En los estándares actualmente existentes se considera principalmente la información sobre la última modificación acontecida en los datos clínicos; quién la realizó, en qué momento, por qué, etc. Es decir, básicamente información de versiones pero no sobre los accesos de consulta que se hayan podido realizar a la misma. La existencia de componentes específicos para el registro de toda actividad de interés acontecida en los sistemas puede facilitar que toda esta información de versiones, responsables… no sea almacenada junto a la clínica y sea gestionada de forma independiente. Para conseguir

esto es imprescindible una identificación inequívoca de todos los recursos sometidos a este control ya que, como se mostró en la vista de información, estos datos de auditoría están vinculados a aquellos a los que se refieren y será necesario manejar adecuadamente estos vínculos.

En la cuarta parte del estándar 13606 se especifica que los detalles de interacción con el sistema de historia clínica deben ser registrados con propósitos de auditoría, pero que la forma en la que estos registros se realicen será específica en cada sistema en cumplimiento de la legislación aplicable en cada dominio y de las prácticas locales. Esto pone de manifiesto que en el caso de organizaciones interregionales o internacionales, estas tareas son críticas y difíciles de ejecutar sin una correcta estructuración de los componentes implicados. Fundamentalmente los componentes de registro de actividad deberían facilitar, de forma normalizada y segura, la consulta externa de registros, principalmente para facilitar las tareas de conocer quién, cuándo y para qué se accede a información de seguridad crítica.

6.5.10. Componentes de comunicaciones

Como se indicó anteriormente, este grupo de actividades ha sido incluido en esta Tesis debido a la importancia de considerar los distintos modelos de interacción que se pueden encontrar entre los numerosos componentes implicados en una organización sanitaria. De manera que tanto la comunicación con sistemas externos como la definición de protocolos de comunicación entre componentes o de distintos mecanismos de acceso a la arquitectura pertenecerían a este conjunto de tareas.

- **Componentes para las comunicaciones externas:** actúan como pasarela o puente entre la arquitectura y elementos externos a la misma. La forma de comunicación más habitual será el intercambio de mensajes, aunque pueden existir otros procedimientos. La tarea de estos componentes será básicamente crear el mensaje en el formato adecuado al interlocutor y utilizar la tecnología apropiada para enviar dicho mensaje, lo cual puede requerir el uso de protocolos de comunicación distintos a los utilizados en la plataforma de la organización. Un caso práctico sería un componente que enviase al usuario un aviso de cita a través del fax. El componente es avisado por alguno de los componentes de gestión de contactos, edita el documento de cita y lo envía al número de fax del usuario. Al recibir mensajes desde fuera, estos componentes extraen la

información útil de los mismos y la convierten a los formatos de la federación transmitiéndola a los componentes adecuados para que lleven a cabo su gestión. Un buen referente para la definición estos componentes del modelo es el EN13606-5, donde se definen los formatos de los mensajes para el intercambio de historias clínicas, aunque hay siempre que considerar que en una arquitectura sanitaria no toda la información será referida a la historia clínica y por tanto hay que extender este modelo.

- **Componentes para las comunicaciones dentro de la organización:** en la vista computacional de HISA se deben normalizar las arquitecturas básicas de los modelos de comunicación más típicos en las organizaciones sanitarias, identificando los componentes mínimos que deben considerarse para dar soporte a las comunicaciones implicadas en estos modelos. Estas arquitecturas deben ser definidas con la participación de expertos en cada uno de los escenarios a normalizar. A continuación se nombran algunos de los modelos más habituales y se realizan algunas consideraciones a tener en cuenta para el diseño de los componentes que se deben incluir.
    - **Modelo de atención hospitalaria:** la red de acceso a la plataforma será, generalmente, la propia red corporativa, ya que el acceso se realiza desde el mismo local de la organización, el hospital. El paciente no suele interaccionar directamente con el sistema sino que es atendido por profesionales que son los que accederán a la plataforma para realizar las consultas o modificaciones apropiadas. En el desarrollo de pruebas diagnósticas o terapéuticas parte de estos registros de información pueden hacerse directamente desde la maquinaria utilizada, sin la intervención directa del profesional sanitario. Cuando es una persona la encargada de interaccionar con el sistema utilizando alguna de las aplicaciones disponibles el terminal puede ser desde un PC de sobremesa a una PDA o un simple busca.
    - **Modelo de atención ambulatoria:** en esta ocasión también están presentes en el mismo emplazamiento físico el paciente y el profesional que lo atiende siendo este último el que interacciona con el sistema. El acceso también es en un local de la organización y por tanto se hará a través de la red corporativa. La mayor actividad ambulatoria es de

atención en consulta, por lo que el terminal más utilizado para gestionar la información a registrar o consultar será el PC o similar, aunque también pueden realizarse algunos tipos de pruebas diagnósticas de forma ambulatoria para las que se utilizan otros terminales más específicos. Por ejemplo, los resultados de pruebas de laboratorio pueden ser introducidos directamente por las máquinas que realizan estas pruebas.

- **Modelo de teleconsulta:** En este caso se podría distinguir entre consultas con el paciente e interconsultas; en cualquiera de los dos supuestos se utiliza la comunicación interpersonal para lograr un diagnóstico o encontrar la terapia adecuada. En el primer caso en uno de los extremos de la comunicación está el paciente, que puede estar en su domicilio particular, (pudiendo el acceso realizarse a través de redes de acceso residenciales comunes, como la RTC, ADSL o una red de cable), o en un local de la organización acondicionado a tal efecto (el acceso podría hacerse a través de la propia red corporativa). También se podría considerar un modelo de teleasistencia como una subclase del modelo de teleconsulta. En el segundo caso la consulta se realiza entre profesionales y el paciente no tiene por qué estar presente. La teleconsulta puede ser en tiempo real o en diferido. Puede necesitarse una conexión con retardos, variación del retardo y ancho de banda determinados, por ejemplo para establecer una videoconferencia, o simplemente un servicio tipo “mejor esfuerzo”, para enviar una consulta por correo electrónico. Los componentes implicados en la teleconsulta pueden utilizar la plataforma sanitaria para acceder a información de interés en la realización del diagnóstico o en la elección de terapia, o para almacenar información recogida durante la misma.
- **Modelo de telemonitorización:** En este modelo el paciente porta sensores que informan a los componentes de la arquitectura adecuados sobre la situación del mismo. Sería destacable la necesidad de componentes que identifiquen posibles causas de alarmas, por ejemplo a partir de un análisis de señales, y componentes que se encarguen de gestionar adecuadamente dichas alarmas. La interacción de estos

componentes con otros de la arquitectura que aportaran más conocimiento sobre el paciente sería crucial para optimizar la respuesta ante posibles alarmas. Si la telemonitorización se efectúa en el domicilio del paciente, la red de acceso puede ser cualquiera de las redes de acceso residencial existentes; el uso de tecnologías inalámbricas en la red terminal permite un mayor movimiento del paciente en su domicilio. En el caso de sistemas de telemonitorización móviles, las posibilidades quedan prácticamente reducidas al uso de redes públicas de telecomunicación móvil cuando se pretende una cobertura suficientemente extensa. Hay que tener en cuenta que el ancho de banda necesario no suele ser muy grande ya que la información puede ser procesada antes de ser enviada. La telemonitorización asistida permitiría la interacción del usuario con los profesionales que atienden el sistema a través de terminales vocales o PCs.

6.5.11. Conclusiones

Aunque se han repasado las principales actividades en una organización sanitaria, es necesario considerar que éstas no son las únicas que pueden existir en la misma. En HISA no están contempladas, y también han quedado fuera de los objetivos principales de este trabajo, tareas tan importantes como la formación de profesionales (tanto de grado como de postgrado), o la investigación clínica. Una arquitectura diseñada formalmente siguiendo el modelo ODP facilitará la incorporación de los componentes necesarios para estas nuevas tareas y la reutilización eficiente de los ya existentes.

6.6. Aproximaciones a las vistas de ingeniería y tecnológica

Según la norma ODP, para completar la especificación de un sistema distribuido sólo quedarían las vistas de ingeniería y tecnología. La primera da un punto de vista sobre el sistema y su entorno destacando los mecanismos y funciones requeridas para soportar la interacción distribuida entre objetos del sistema. En la segunda se proporciona una vista del sistema y su entorno donde se presentan las características particulares de la tecnología elegida para la implementación del sistema.

HISA especifica que al ser estas dos vistas propias de la tecnología subyacente no deben formar parte del estándar. Sin embargo, hay una consideración que se debe realizar al

respecto, hay que tener en cuenta que, como se ha dicho anteriormente, muchas de las características de un sistema sanitario se pueden encontrar en otros dominios. Las nuevas plataformas de computación distribuida pueden resolver algunas de las tareas identificadas en la vista de ingeniería y, por tanto, no sería necesaria la normalización específica dentro de HISA, ya que la plataforma sanitaria es una especialización que se superpone a alguna plataforma de integración genérica, como se vio en la figura 11.

De hecho muchas necesidades no se han nombrado en las vistas de alto nivel aunque son imprescindibles para el correcto funcionamiento de la plataforma. Por ejemplo la identificación de componentes, la publicación y localización de los mismos, los mecanismos de establecimiento de vínculos, de invocación de interfaces, etc, todas se consideran parte de la plataforma subyacente.

Es más, cuando el desarrollo de las plataformas de integración basadas en la descripción semántica del modelo de información y funcional llegue a buen puerto, el desarrollo de una arquitectura sanitaria que permita reutilizar componentes para ofrecer aplicaciones de altas prestaciones a los usuarios de la organización no necesitará el gran esfuerzo de normalización que se está realizando actualmente. Hasta ese momento la normalización propuesta por HISA introduciendo, como se propone en esta Tesis, los avances realizados en la gestión de ontologías para dinamizar el modelo de información y funcional es una excelente herramienta para lograr la interoperatividad de sistemas heterogéneos y que facilita la integración según un modelo federado.

Capítulo 7. Validación: gestión demográfica

7.1. Paradigma estratégico

En las actividades del sujeto de cuidado se incluyen aquellas para la gestión de pacientes cuyos objetivos fundamentales son la identificación inequívoca del mismo y la gestión de sus datos personales. Aunque en HISA se distingue entre tareas de gestión de usuarios y de pacientes, muchos componentes pueden ser comunes a ambas, de hecho tanto pacientes como usuarios son habitualmente personas, aunque con distintos papeles en la organización. De este modo las actividades de gestión de personas están muy relacionadas con varios grupos de actividades identificados en la vista de empresa, de manera que los componentes que las soportan podrían ser ampliamente reutilizados si se hace un diseño eficiente.

En esta Tesis Doctoral se presentan las vistas de empresa, información y computacional para las tareas de identificación de personas y de gestión de datos demográficos. Los objetivos principales de los componentes que soportan estas actividades en una arquitectura sanitaria serán:

- Registro de personas.
- Identificación de personas.
- Recuperación de datos demográficos o personales.
- Actualización de datos demográficos o personales.

Estos componentes serán utilizados, entre otros, por los agentes dedicados a actividades relativas a la gestión de usuarios del sistema, de sus credenciales y roles. También son fundamentales en las relacionadas con la gestión de pacientes y sus contactos con la organización. Serán piezas clave en actividades más complejas dentro de la federación, como las de gestión de historia clínica o administrativa.

Uno de los conceptos más importantes en cualquier sistema de gestión de personas es el de identificador unívoco de persona. Cuando una persona se registra en un sistema se le asigna un identificador único que se va a utilizar como índice o clave cada vez que se almacene información relativa a esa persona, sea cual sea la naturaleza de la misma, y

se pretenda mantener ese vínculo información-persona. Así, por ejemplo, en un sistema de historia clínica una vez conocido el identificador de la persona se pueden recuperar las observaciones de interés relativas a la misma.

Otro concepto fundamental es el de atributo. Un atributo es una característica determinada acerca de las personas registradas, es decir los datos personales que se van a manejar en un determinado sistema. Por lo general se incluye información de identificación, como el nombre y apellidos o el DNI, de localización, como la dirección postal o el teléfono, o de otra naturaleza como el sexo, la religión o la fecha de nacimiento. Cuando se trata de pacientes, en la mayoría de los sistemas también se reflejan datos de carácter biológico o clínico que son especialmente importantes para su atención sanitaria, como la existencia de alergias o su grupo sanguíneo. Estos datos pueden considerarse observaciones clínicas, y su gestión se haría utilizando los componentes de información clínica. Sólo aquellos que puedan aportar información para la identificación unívoca del paciente van a ser tratados como atributos de persona por los componentes demográficos.

Las actividades de gestión de personas y datos demográficos deben estar asignadas a componentes especializados. Como muestra la figura 46, la normalización en el seno de la arquitectura sanitaria debe incluir la vista de empresa, que analizaría las funciones propias de estas tareas, la de información, que detallaría los conceptos básicos manejados, traducidos en clases y tipos, y la computacional, que daría las interfaces básicas para manejar los objetos del modelo de información y las que realizan prestaciones más avanzadas. Los nuevos sistemas de información que se desarrollen deben estar basados en estos modelos de datos e interfaces para facilitar la incorporación de dominios a una organización sanitaria siguiendo el modelo de federación. Sin embargo HISA debe proveer de mecanismos que faciliten la incorporación de dominios en los que el registro de personas y la gestión de datos demográficos se realicen de modo distinto. Cosa que ocurrirá especialmente si no se han realizado desarrollos basados en estándares o si los estándares elegidos han sido distintos en cada uno. Aunque en este supuesto el esfuerzo a realizar para integrar el sistema en la federación será mayor, éste puede reducirse considerablemente si se planifican soluciones abiertas de integración.

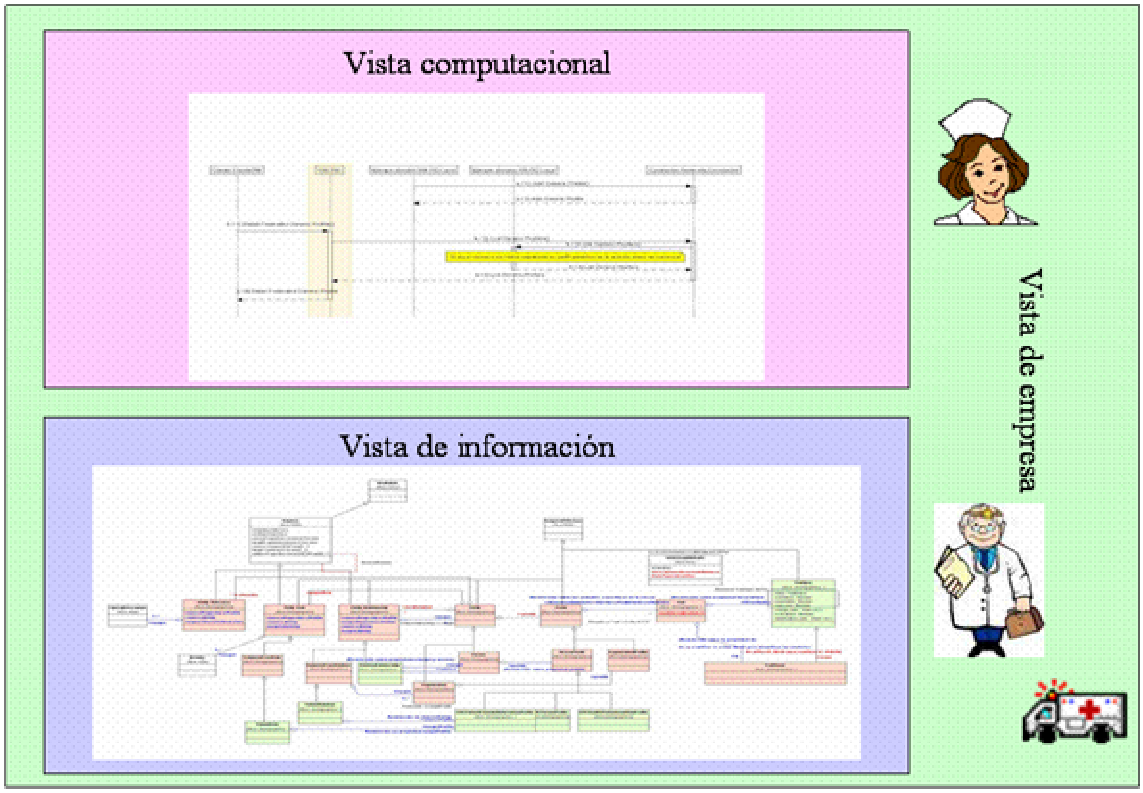


Figura 46. Vistas de la arquitectura

La normalización de HISA no está concluida actualmente, de manera que esta Tesis propone soluciones basada en lo ya descrito en las primeras versiones de HISA y en otros estándares anteriormente desarrollados. Además se presta especial atención a la problemática en entornos federados, es decir se proponen soluciones a la integración de dominios autónomos dentro de la arquitectura sanitaria, siempre en lo referido a la identificación de personas y la gestión de datos demográficos.

En cuanto a la federación de dominios administrativos, y en lo que a las tareas de gestión de personas se refiere, pueden identificarse principalmente los siguientes obstáculos:

- La utilización de distintos modelos de información, y por tanto distintos conceptos y atributos de persona y/o distinta forma de nombrarlos, en cada uno de los dominios a integrar.
- El uso de distintas interfaces para la gestión de datos demográficos para cada sistema.

- La no existencia de un identificador de persona único compartido por todos los sistemas a integrar.
- La existencia de inconsistencias de valores de atributos en distintos sistemas de la federación.

La normalización de las vistas de información y computacional solucionan, en parte, el primer y segundo problema. Permiten tener una semántica común de atributos de persona y un modo de acceso normalizado a esta información. Para llevar a cabo la integración de sistemas no desarrollados conforme al estándar sería necesaria la implementación de componentes mediadores entre cada sistema local y el resto de la federación. Estos mediadores cumplirían tanto con el modelo de datos como con el computacional, ocultando a la federación los detalles locales.

Aunque HISA aconseja la utilización del mismo identificador de persona en todos los dominios de la organización, es habitual que éste no exista, sobre todo cuando se mantiene la autonomía de los sistemas y cuando éstos se han creado previamente a la integración en la organización. Es evidente que la existencia de identificadores administrativos como pueden ser el DNI o el NSS, es decir identificadores asignados a las personas por alguna entidad oficial, los convierte en claros candidatos a ser utilizados como identificadores de persona, únicos y universales. Sin embargo ciertos problemas como la posible duplicidad de éstos, los frecuentes errores tipográficos o el hecho de que algunas personas no los tengan asignados, impiden su uso como identificador único y esto ha dado lugar a que en la mayoría de dominios el identificador de persona se asigne y se gestione localmente. Por otro lado una arquitectura que plantee una alta escalabilidad no puede depender totalmente de identificadores asignados por una organización ya que si se integran sistemas de distintas administraciones estos identificadores pueden no concordar. En este caso serán necesarios otros mecanismos que permitan la correcta identificación de la persona, y en especial de pacientes, en cada uno de los dominios que se integren en la federación.

La información demográfica es particularmente sensible a los problemas de reconciliación de entidad, ya que será habitual que la misma persona, en especial el paciente, esté registrada en varios dominios y que se almacene la misma información demográfica sobre ella. Es decir existen distintas versiones de la misma entidad, en este caso persona, del mundo real. Por otro lado al mantener la autonomía de cada dominio

no es de extrañar que los valores del mismo atributo se actualicen de forma completamente independiente en cada dominio y por ello, a menudo, no coinciden. Las dos causas fundamentales de conflicto en los valores de un atributo correspondiente a la misma persona pero registrado en sistemas distintos son:

- **Actualización desincronizada del atributo:** si el valor del atributo ha cambiado pero no todos los sistemas tienen constancia de ello se tendrán múltiples valores para el mismo atributo. Guardar registro de la fecha de modificación de un atributo y de los periodos de validez para el mismo puede ayudar en la resolución de este tipo de conflictos.
- **Discrepancia en la recogida de valores:** que el registro del mismo valor de un atributo se haga de forma distinta en los sistemas. Dos causas frecuentes son:
    - La posibilidad de utilizar varias formas para representar el mismo atributo, por ejemplo usar sólo mayúsculas o que algunos apellidos compuestos se puedan representar usando guiones o sin ellos, si el nombre es compuesto usar sólo uno de ellos, etc.
    - Que se cometan errores tipográficos, que serán muy frecuentes considerando que en muchas ocasiones la información se introduce a través de interfaces de usuario
    - El manejo de datos siempre canonizados según determinado criterio y basar la identificación en varios parámetros cruzando los resultados pueden ayudar a resolver estos conflictos.

Pueden existir distintos modelos de gestión demográfica en una organización sanitaria que integre varios sistemas de información:

- **Centralizado:** toda la información de personas se recoge en un único sistema, como se muestra en la figura 47. Este modelo es fácil de gestionar. Cuando un dominio se incluye en la federación, la información de sus personas se traslada al sistema central de identificación de personas y se externalizan todas las tareas de gestión demográfica. Durante el registro en el sistema central la primera tarea será la identificación de la persona en el mismo, ya que ésta podría estar previamente registrada. En caso de que la persona ya estuviera registrada habría que cotejar los atributos de la misma y actualizarlos si fuera necesario. Hay que

tener en cuenta que al registrarse en el sistema central se le puede asignar un nuevo identificador, y en ese caso habría dos modos de actuación: o bien el identificador único que la persona tenía dentro del sistema local se modifica también, lo que supondrá la actualización de todas las apariciones del mismo, o bien es el sistema central el que se encarga de hacer la correspondencia entre identificadores locales y el identificador único en la federación. El mayor problema de este modelo es la pérdida de autonomía en cuanto a la gestión de datos demográficos y la fuerte dependencia de un sistema central que, entre otras cosas, hace la arquitectura muy sensible a fallos en el mismo y poco escalable.

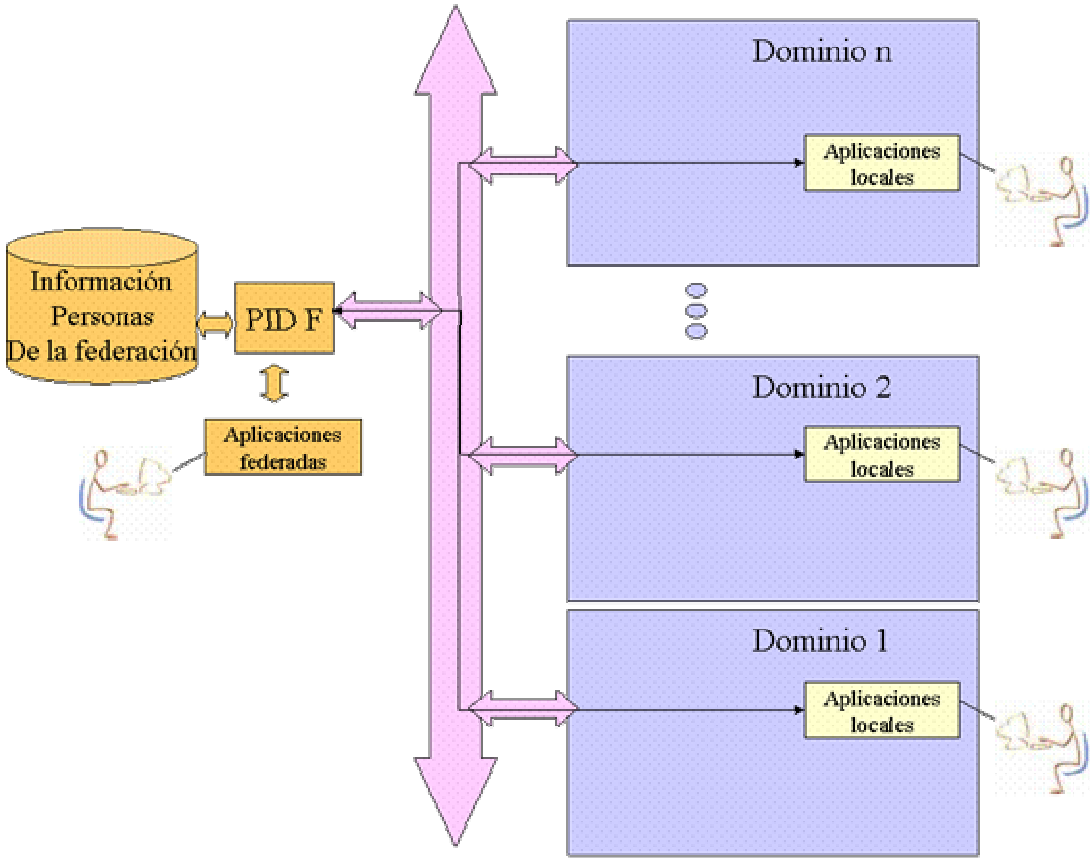


Figura 47. Centralización de la gestión demográfica

- **Semi-centralizado:** los registros se siguen haciendo en cada sistema de forma independiente, asignando localmente los identificadores de persona, pero la información se duplica en un único sistema central, como se muestra en la figura 48. Éste se encarga de relacionar los dominios encontrando los identificadores para la misma persona en los distintos sistemas. Cada dominio sigue funcionando como antes de entrar en la organización, por lo que se mantiene

totalmente la autonomía, aunque debe informar al sistema central de las actualizaciones de la información demográfica de cada persona para mantener la consistencia. Sólo hacen uso del sistema de correlación los componentes federados que necesitan información sobre la misma persona pero proveniente de distintos dominios. En este caso el sistema centralizado se encarga de resolver los conflictos de entidad y de proveer un identificador único para referirse a esa persona en todos los dominios de la organización. El mayor problema de este modelo es la gran capacidad de almacenamiento que se necesita en el sistema central, ya que debe guardar información de todas las personas registradas en cada uno de los dominios. Es decir que tendrá almacenadas varias instancias de la misma persona, una por cada sistema en el que esté registrada. Se sigue dependiendo del nodo central pero esta dependencia es menos crítica, al menos en todas las tareas no federadas, en las que ni siquiera se utiliza salvo para informar de los cambios realizados en los valores de los atributos de la persona. Este modelo es el utilizado por el servicio PID de CORBAmed para realizar la correlación de dominios.

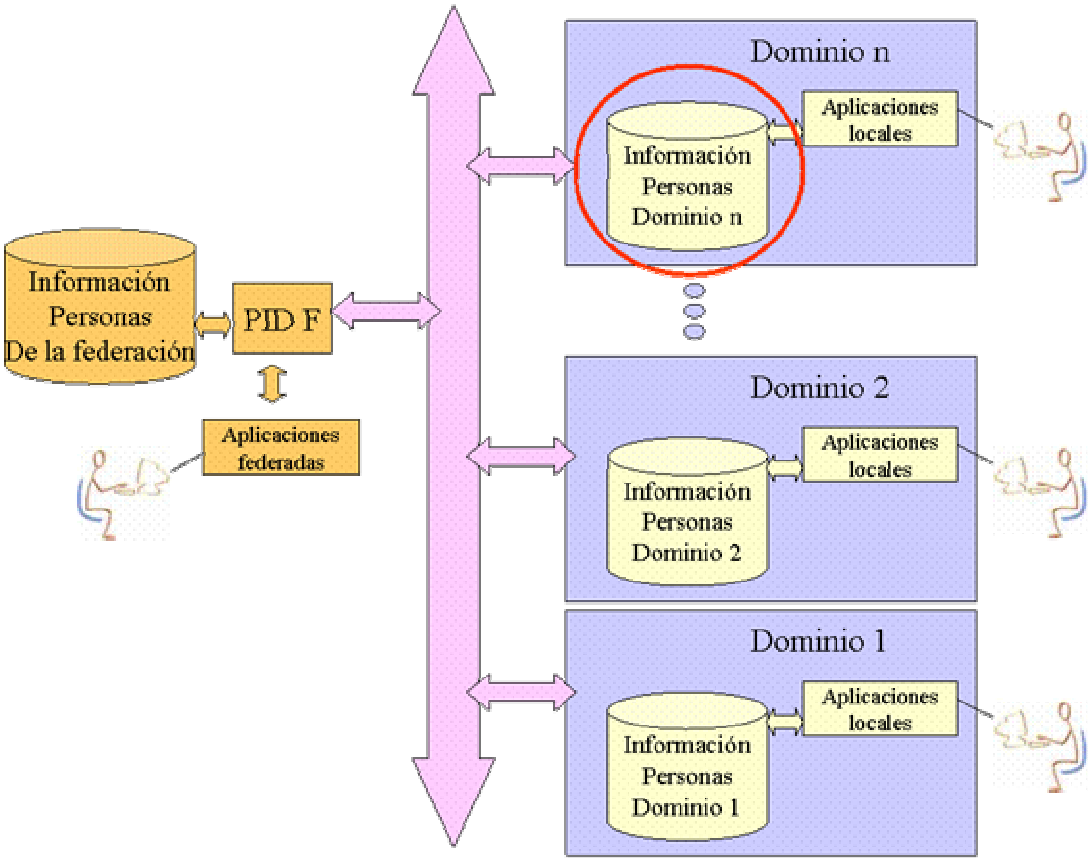


Figura 48. Gestión semi-centralizada de la información demográfica

- **Federado:** el registro del paciente se hace de forma independiente en cada sistema de la organización, sin que exista ninguna centralización de la información de personas, por lo que los identificadores y datos demográficos sobre la misma persona pueden diferir en cada sistema, como se muestra en la figura 49. Lo que sí se centraliza es información sobre las claves de identificación en cada dominio, es decir, sobre los nombres de los atributos que sirven de identificadores unívocos para cualquier persona en un sistema. En otras palabras, se centraliza la información sobre las claves únicas que pueden resolver la identidad de una persona. Esta información es única para cada sistema, no para cada persona, por lo que el volumen de información centralizada es mucho menor que en el modelo anterior. En todas las tareas no federadas se trabaja como si no se perteneciera a una federación, de forma totalmente autónoma. Sólo se informa al sistema central si se cambia algo de la estructura interna del sistema de información que implique un cambio en las claves de identificación, aunque estas modificaciones en el esquema de los sistemas no son muy habituales. En los componentes federados se recurre al sistema central para conocer el modo de identificar a una persona en cada uno de los dominios de la organización. Los conflictos de entidad se resuelven "al vuelo", en el momento en el que un componente federado realiza la consulta de datos demográficos. Durante esta consulta y todo lo que dure la sesión correspondiente, se puede asignar y mantener un identificador de persona sobre toda la federación y las correspondencias de este identificador temporal a los identificadores en cada dominio. Esto permitirá recuperar cualquier información relativa a la persona en cualquiera de los sistemas.

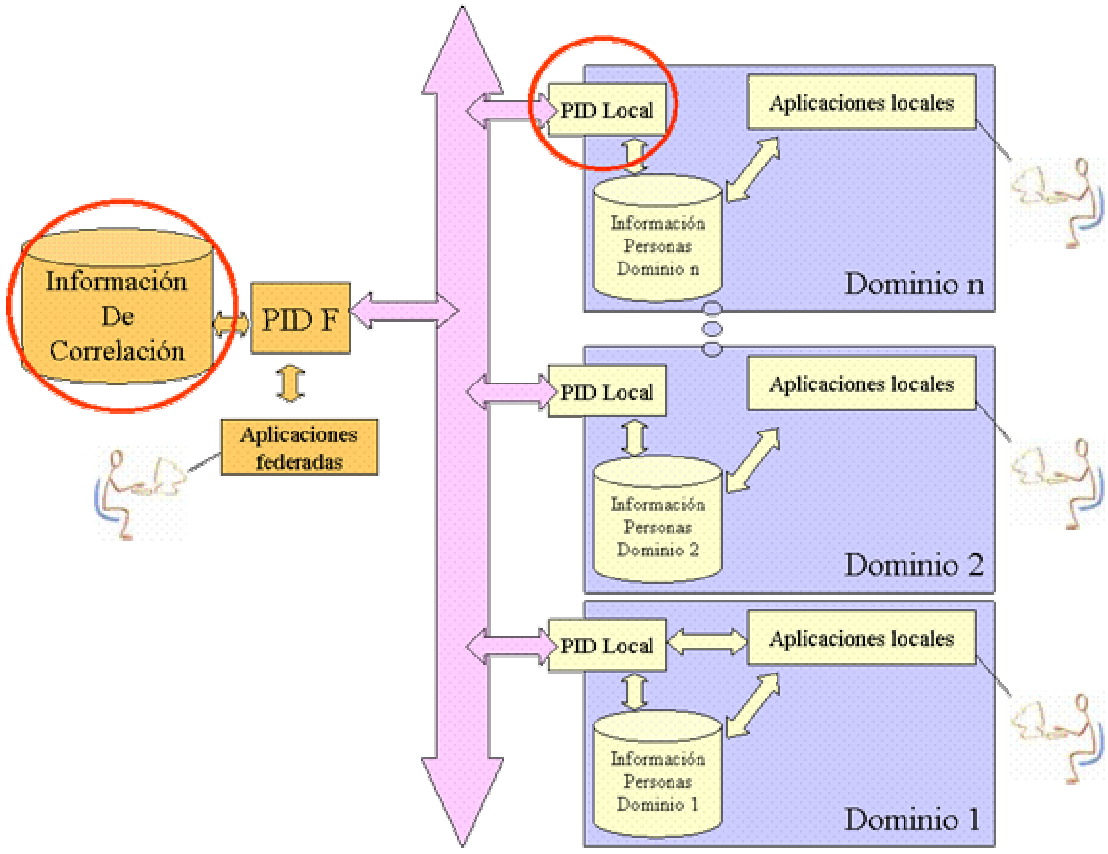


Figura 49. Federación de la gestión demográfica

- **Colaborativo:** es un sistema más complejo en el que se mantienen los registros locales, como se muestra en la figura 50, pero se asegura la consistencia de la información demográfica referida a la misma persona en cada momento. Cada sistema debe conocer la información que maneja el resto de los sistemas de la federación y cuando se modifique localmente información demográfica referida a una persona informará de los cambios a todos aquellos sistemas que tengan constancia de la misma información para que también la actualicen. La información demográfica por tanto estará duplicada en varios dominios aunque no existirá el problema de inconsistencia de instancias. Los agentes demográficos de cada sistema son complejos, ya que deben resolver los problemas de identificación unívoca dentro de la federación. Las aplicaciones federadas podrán utilizar los servicios demográficos de cualquier sistema que guarde la información de interés acerca de la persona, ya que es seguro que ésta estará actualizada. Este modelo es parecido al centralizado en el sentido de que se ha perdido la total autonomía, ya que es necesario diseñar para cada sistema

nuevos componentes de gestión demográfica, aunque lo aventaja en robustez dado que no hay dependencia de un nodo central. La gran dificultad de este modelo es la complejidad que entraña el desarrollo de mecanismos que permitan que cada sistema tenga conocimiento acerca de toda la federación en cualquier instante. Por otro lado las tareas de actualización síncrona de información no resultan sencillas.

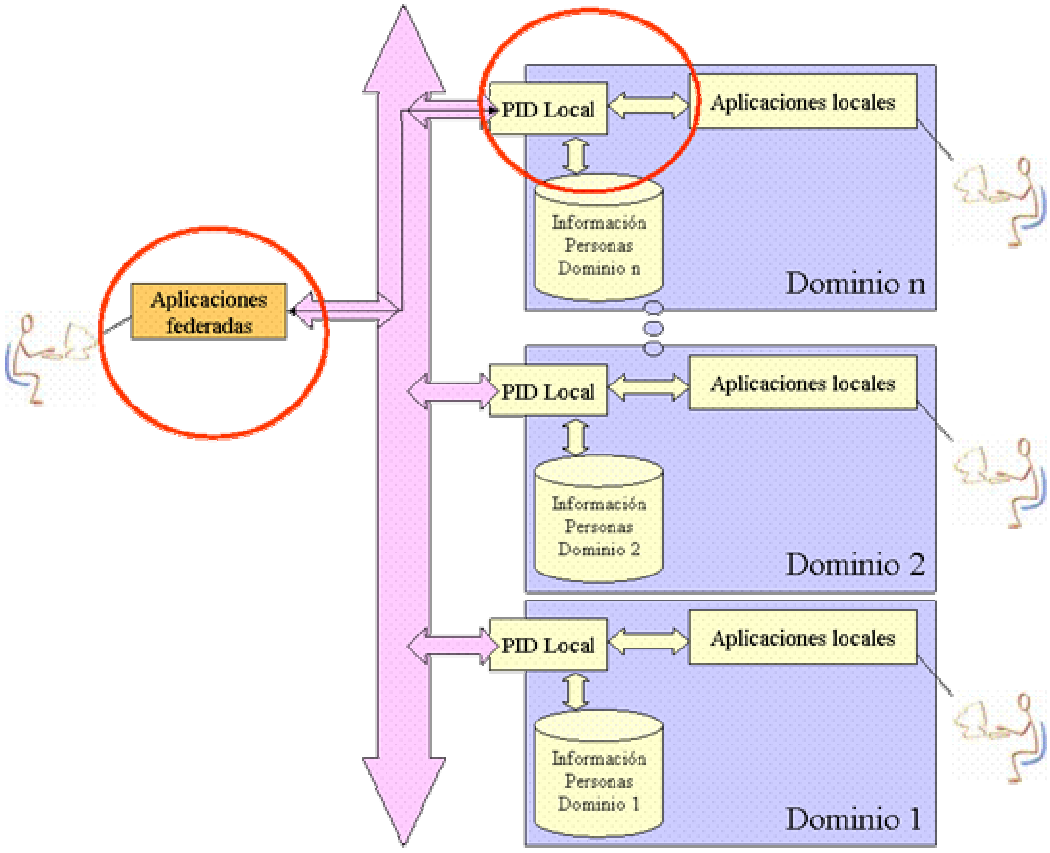


Figura 50. Modelo colaborativo para la gestión de información demográfica

El modelo desarrollado en esta Tesis es el federado, ya que se considera el más escalable, eficiente y cercano a la realidad existente en las organizaciones sanitarias.

Como se puede apreciar, este trabajo no cubre todo el conjunto de actividades del sujeto de cuidado, dado que no se van a considerar las tareas de gestión de contactos. Por otro lado, abordar la gestión de personas y no específicamente de pacientes supondrá la reutilización de algunos componentes en otros grupos de actividades; en especial en las tareas de gestión de usuarios y seguridad.

Una vez presentado el planteamiento general de los componentes de gestión demográfica se detallan las vistas de empresa, de información y computacional.

7.2. Vista de empresa

7.2.1. Introducción

Las actividades de gestión de personas forman parte de las actividades del sujeto de cuidado y de las de gestión de usuarios de HISA, aunque los componentes que las soportan van a ser reutilizados en muchas otras actividades de la organización sanitaria.

Las tareas básicas identificadas dentro de estas actividades son:

- **Registro:** consiste en introducir una nueva persona en el sistema. Esto incluye la asignación de al menos un identificador unívoco, el identificador de persona o clave primaria, y la recopilación y almacenamiento de su información personal. Esta información es independiente de los futuros contactos que esta persona tenga con la organización y entre otros puede contener datos personales o demográficos (nombre, identificadores, fecha de nacimiento…), relaciones con otras personas registradas, información de seguros o de atención primaria, papel o papeles de la persona dentro de la organización. En la mayoría de los sistemas también se recogen durante el registro algunos datos genéricos de salud (grupo sanguíneo, factores de riesgo, alergias, antecedentes personales…), especialmente cuando la persona tiene el papel de paciente en la organización. Respecto a estos datos clínicos o biológicos sencillos, fundamentales en la atención al paciente, podría hablarse de un conjunto mínimo de datos personales de interés sanitario. Al ser información de tipo clínico podrían considerarse como observaciones y por tanto la gestión de la misma se haría con los componentes que dan soporte a las actividades de información clínica. Sin embargo, dado que las características de la gestión de este conjunto mínimo de información son muy próximas a las de los datos demográficos (muy estables en el tiempo, duplicado en muchos sistemas, información que interesa consultar siempre, etc) en algunos sistemas se tratan como atributos de la persona y se utilizan para su gestión los mismos componentes que para la información demográfica. En este trabajo de entre los datos biológicos recogidos en el

registro únicamente se consideran atributos personales aquellos que puedan ayudar en la identificación unívoca de la persona.

- **Identificación:** la finalidad es encontrar el identificador unívoco de la persona de interés. Generalmente este identificador será el índice para recuperar el resto de la información relativa a la persona, tanto sus atributos personales como clínicos o de cualquier otra naturaleza. Esta identificación se realiza a partir de la información conocida de la persona, es decir de sus atributos. Un conjunto de atributos se conoce como un **`perfil`**. Cuanto más se restrinja la búsqueda, introduciendo más atributos conocidos, menos candidatos se obtendrán y se estará más cerca de obtener el identificador de persona deseado.
- **Recuperación de atributos:** una vez conocido el identificador se puede recuperar la información personal almacenada en el sistema. La consulta se hará a partir del identificador, indicando los atributos que se desea conocer sobre el sujeto y obteniendo por tanto un perfil de dicha persona.
- **Actualización de atributos:** algunos datos personales pueden cambiar con el tiempo de forma que deben existir mecanismos para la modificación de los atributos de persona.

Se pueden también identificar algunas tareas más complejas como:

- **Gestión de identificadores:** el registro se realiza cuando se quiere introducir una nueva persona, después de comprobarse que esa persona no estaba ya registrada. El registro incluye la asignación de un identificador unívoco pero en esta asignación pueden surgir algunos problemas. Por ejemplo, dado que la búsqueda de personas se hace a través de atributos conocidos sobre la misma, si éstos no son correctos pueden ocurrir errores en la identificación y por tanto puede no encontrarse una persona que realmente sí está registrada. Esto puede dar lugar al registro de la misma persona varias veces, es decir la misma entidad real está representada por objetos de información distintos dentro del mismo sistema. Si posteriormente se comprueba el error puede solucionarse fusionando los identificadores, pero hay que tener en cuenta que para el resto de información almacenada sobre la persona se podría haber utilizado cualquiera de los identificadores como índice, por lo que hay que mantener ambas referencias o modificar todas las entradas. Esta última opción impediría volver a separar los

dos identificadores en caso de haber cometido un error al fusionarlos. Para contemplar toda la casuística posible los identificadores tendrán distintos estados y se necesitan componentes que manejen la transición entre estados.

- **Recuperación de perfiles personalizados:** algunos componentes pueden ofrecer los datos personales de un individuo de forma contextualizada, es decir, que devuelven justo los atributos que se necesitan en un momento determinado, sin tener que especificar qué atributos se desean.

A continuación se desglosan estas actividades en tareas más específicas para facilitar la descomposición en componentes que se presentará en la vista computacional. Se comienza identificando las actividades que afectan únicamente a cada sistema de la arquitectura, para luego pasar a identificar las que se refieren a toda la federación. Para llevar a cabo estas actividades se reutilizarán los componentes desarrollados para las primeras y se tendrán que resolver los problemas inherentes al modelo federado.

7.2.2. Actividades de gestión de personas sobre sistemas autónomos

Cada uno de los sistemas a integrar en la federación, y que incluya entre sus actividades la gestión de personas, debe realizar las tareas presentadas en este punto. Estas son las propias de la gestión de personas, identificadores y datos personales dentro de uno de los dominios adscritos a la federación, sin considerar los demás. Es evidente que el manejo de datos personales es especialmente crítico en cuanto a seguridad. El cumplimiento de las leyes vigentes en cuanto a protección de datos personales y de las políticas de acceso establecidas en el sistema debe quedar garantizado. Esto implica la existencia de vínculos entre los componentes de gestión de personas y los pertenecientes al grupo de actividad de seguridad y gestión de recursos (usuarios, credenciales, control de acceso, auditoría, etc). Las tareas más importantes en este grupo de actividades son:

- **Tareas de identificación:** a partir de los atributos de una persona se obtiene el identificador único de la misma. La combinación de atributos utilizada para la búsqueda es determinante en el número de candidatos que localice el componente. Pueden existir combinaciones de atributos que, por diseño del sistema, aseguren que el número de candidatos devueltos es siempre como máximo 1. Por ejemplo si el desarrollador del sistema ha impuesto que no pueden existir dos personas que tengan el mismo DNI, nombre, primer apellido

y segundo apellido esta combinación de atributos garantiza una identificación inequívoca. Al conjunto de nombres de atributos que componen una clave única se le denomina **`perfil genérico`**. En el siguiente apartado se explican en más detalle las tareas en entornos federados y se aclarará el uso de los perfiles genéricos. Hay que tener en cuenta que solamente se permite el uso de estas prestaciones a los componentes autorizados y que actúen en nombre de usuarios con derecho a la consulta de los datos de la persona buscada en cuestión. Las respuestas a consultas nunca deben vulnerar el derecho a la privacidad de la persona, ni siquiera las respuestas negativas [41]. Así, por ejemplo, si la persona es encontrada pero el usuario que pretende hacer la consulta no tiene derecho a conocer los datos personales del mismo la respuesta no debería permitirle deducir que realmente esa persona está registrada en el sistema.

- **Tarea de registro:** permite introducir una nueva persona en el sistema. Asignará un nuevo identificador único y almacenará los atributos del nuevo miembro, apoyándose para ello en los componentes de gestión de atributos. Solamente componentes y usuarios autorizados podrán realizar este registro.
- **Tareas de consulta de perfiles:** a partir del identificador de una persona se permite la recuperación de los atributos de la misma. Solamente se permite la consulta de los atributos que el usuario esté autorizado a conocer. Existirán consultas básicas y otras más complejas que faciliten la recuperación de perfiles contextualizados.
- **Tareas de gestión de atributos:** a partir del identificador de una persona se permite la modificación, borrado o inserción de los atributos de la misma a componentes autorizados que actúen en nombre de usuarios con derecho a la gestión de esos atributos de la persona.
- **Tareas de gestión de identificadores:** realizan tareas de gestión de identificadores como el cambio de estado del identificador, la fusión de identificadores o la posterior separación.

Todos los componentes que lleven a cabo este conjunto de actividades deben ser conformes al modelo de información presentado en la vista de información y a la interfaz funcional presentada en la vista computacional. El modelo de información describirá los atributos de persona, las relaciones entre los mismos y los tipos de datos

que los soportan. El modelo funcional especificará las interfaces de estos componentes y los posibles vínculos entre ellos.

No hay que olvidar que en una arquitectura basada en la federación de sistemas es necesario respetar la autonomía de los sistemas desarrollados previamente a la integración en la plataforma, ofreciendo al mismo tiempo facilidades para realizar esta integración. Considerando esto y el modelo de arquitectura presentado en el capítulo anterior existen dos filosofías para enfrentar el problema de la integración:

La primera posibilidad es considerar el modelo de información estático y una especificación funcional rigurosa. Esta solución por tanto es perfectamente compatible con la estrategia actual de desarrollo de HISA. En este caso, para los sistemas que no cumplan con esta especificación deben realizarse tareas de adaptación que podrían dividirse en:

- **Tareas de adaptación semántica:** facilitan la traducción de los conceptos locales a los especificados en la vista de información y viceversa. Se pueden utilizar componentes del grupo de actividades de gestión semántica si estos están ya desarrollados en la arquitectura. La adaptación semántica incluye la identificación de conceptos equivalentes y de las conversiones de tipos de datos necesarias.
- **Tareas de adaptación funcional:** actúan de mediadores entre el sistema y la federación. Son componentes que interactúan con el sistema local cumpliendo las interfaces definidas en el mismo y presentan a la federación la interfaz normalizada en la vista computacional. Cuando sea necesario, utilizan los componentes de adaptación semántica para resolver los conflictos semánticos.

En el otro extremo se puede modelar el dominio demográfico como una ontología dinámica y extensible, como se presentaba en el capítulo anterior. Esto permitirá que al introducir un nuevo sistema se incluyan los atributos manejados por éste en la ontología federada para que todos los demás puedan también utilizarlos. Los componentes semánticos especializados en la gestión de ontologías se encargarán de hacer las traducciones y equivalencias necesarias. En una arquitectura avanzada de gestión semántica la vista computacional puede ser muy relajada, indicando únicamente la funcionalidad mínima requerida de los componentes y el lenguaje impuesto a los

desarrolladores para describir sus componentes. Si éste fuera el caso, simplemente habría que describir los componentes locales conforme al lenguaje normalizado.

Entre estas dos soluciones extremas existen numerosas posibilidades. Dado el actual estado de las plataformas semánticas, la presentada en segundo lugar es, todavía, demasiado futurista, en especial en lo relativo al relajamiento de la vista computacional. Esto indica que la existencia de componentes de adaptación funcional está, por ahora, garantizada y en cuanto a los de adaptación semántica también existirán, aunque es más fácil comenzar a considerarlos parte de la plataforma subyacente a HISA.

7.2.3. Actividades de gestión de personas sobre la federación

Las actividades de gestión de persona sobre la federación resuelven los problemas relativos a la gestión de personas, identificadores y datos personales pero en esta ocasión actuando sobre distintos dominios de la federación. Al igual que en el caso anterior, la relación con las actividades de seguridad es estrecha, pero ahora esta gestión es más compleja ya que el usuario puede tener distintas credenciales en cada sistema. El uso de componentes de seguridad federados asegura el cumplimiento de las normas de acceso y gestión de recursos dentro de la federación. Algunas de las tareas más importantes de estos componentes son:

- **Tareas de identificación:** a partir de los atributos de una persona obtiene un identificador federado de la misma. Este identificador federado lo identifica unívocamente dentro de toda la arquitectura sanitaria o de un subconjunto de dominios de la misma. Para llevar a cabo esta tarea se apoyarán, entre otras, en las tareas de identificación autónomas, debiendo asegurar que la persona identificada en cada sistema corresponde efectivamente con la persona de interés. La dificultad estriba en que el identificador único de persona es gestionado independientemente en cada dominio, por lo que habría que obtener el identificador de cada uno de ellos y mantener las correspondencias entre éstos y el identificador federado.
- **Tareas de gestión de identificadores:** se refieren al manejo de los identificadores federados, incluyendo tareas como la asignación, establecimiento de correspondencias con identificadores locales o el cambio de estado. El identificador federado puede ser temporal y mantenerse únicamente durante una sesión. Entendiéndose por sesión el periodo de uso de una o varias facilidades de

la federación para la realización de determinadas tareas, que comienza normalmente con el acceso seguro al sistema.

- **Tareas de correlación:** las tareas de correlación implementan la lógica necesaria para la constatación de que varios objetos entidad, y en este caso personas, en distintos dominios de la federación se refieren a la misma entidad del mundo real. El paradigma propuesto en este trabajo para la gestión de identificadores en la federación está basado en dos conceptos fundamentales del dominio de bases de datos.
    - **La clave única:** que es una combinación de atributos (puede ser también sólo uno) que identifica unívocamente a un objeto, de forma que el sistema de gestión de base de datos asegura la no duplicidad de la misma. Podría ser generado automáticamente por el propio sistema de gestión de base de datos
    - **La clave primaria:** que es una de las claves únicas que puede haber en un sistema. En una tabla identifica cada una de las tuplas de la misma.

En un dominio, el concepto de identificador de persona, y el más general de identificador de entidad, es equivalente al de clave primaria. Éste se corresponde con el concepto **`PersonId`** utilizado en el servicio PID de CORBAmed. El concepto de clave única será una combinación de atributos que identifiquen unívocamente una entidad, y en concreto también a una persona, dentro de un dominio. Este conjunto de atributos debe ser indicado por el diseñador del sistema. Hay que apreciar la diferencia de que mientras un perfil (**`profile`**) es un conjunto de nombres de atributos y su valor concreto que identifican de forma unívoca a una persona en un sistema, una clave única sería la secuencia de nombres de atributos, sin especificar el valor concreto, que identifica de forma unívoca a cualquier persona. En este trabajo se utilizará el término perfil genérico (**`generic profile`**) para referirse al conjunto de nombres de atributos que identifican unívocamente a cualquier persona en un sistema (es decir una clave única). Por supuesto un sistema puede tener más de un perfil genérico.

Las tareas de correlación se apoyan en la gestión de perfiles genéricos. Por un lado se publican interfaces para el registro de los perfiles genéricos de cada sistema, y por otro para la consulta de los mismos y la obtención del perfil genérico federado. Este perfil genérico federado es el conjunto de atributos que garantiza la identificación inequívoca

de la persona en todos y cada uno de los sistemas de la federación. Usando este paradigma, los componentes de correlación no van a tener que basarse en una centralización de la información demográfica, como ocurría con los trabajos hasta ahora planteados [29,31,32], sino que únicamente centralizan los perfiles genéricos, es decir, información semántica sobre las claves únicas en cada sistema, reduciendo de este modo la duplicidad, inconsistencia e intercambio de información, y aumentando la seguridad y escalabilidad en el sistema. Es evidente que los perfiles genéricos de cada sistema cambian con mucha menos frecuencia que la información demográfica de las personas registradas.

Aparte de esta gestión de perfiles genéricos, los componentes de correlación también utilizarán las interfaces publicadas por los componentes de gestión de identificadores y de consulta de atributos para llevar a cabo sus tareas. Todos aquellos sistemas en los que no se ha autorizado al usuario la obtención del identificador quedan fuera de la identificación federada. En las tareas de identificación sobre toda la federación puede ser necesario resolver algunas inconsistencias en los valores de atributos en distintos sistemas, como se analizará en la vista computacional.

- **Tareas de registro:** ya que la federación implica la autonomía de los sistemas, en la mayor parte de los casos se van a seguir utilizando las aplicaciones locales en las actividades cotidianas, mientras que las aplicaciones federadas darán un valor añadido que, generalmente, se limite a la consulta y no a la introducción de información. Debido a esto sería difícil encontrar el caso de una aplicación que necesitara registrar a una persona en todos los sistemas a la vez. Si así fuera, para mantener totalmente la autonomía de cada sistema y seguir el modelo de federación pura habría que realizar el registro en cada uno de los sistemas, reutilizando los componentes locales de registro. Existirían componentes que gestionarían este registro múltiple a través de una única interfaz.
- **Tareas de consulta de atributos:** a partir del identificador de una persona se permite la consulta de los atributos de la misma para los que el usuario tenga permiso concedido. Se encargan de la reconciliación de valores en aquellos atributos en los que no haya concordancia.
- **Tareas de gestión de atributos:** a partir del identificador de una persona se permite la modificación, borrado o inserción de los atributos de la misma en

todos los sistemas de la federación al unísono; siempre y cuando exista autorización para ello.

Existen algunos referentes bibliográficos [31,32,136] que manejan el concepto de PMI (Person Master Index) como un agente que agrupa todas estas tareas sobre la federación. En la arquitectura propuesta cada una de las tareas aquí identificadas va a ser llevada a cabo por un componente especializado; como la realización práctica no obliga a que esto sea así, ya que el modelo ODP permite que un mismo componente implemente varias interfaces, este concepto es perfectamente compatible.

7.3. Vista de información

El análisis exhaustivo de las características de la información personal permitirá encontrar los factores que condicionan la gestión de la misma. Por eso es interesante analizar detenidamente estas características.

El primer rasgo característico es la alta duplicidad de esta información, ya que la mayoría de los sistemas registran, en mayor o menor medida, la misma información demográfica. Sin embargo las observaciones clínicas, por ejemplo, no suelen duplicarse sino que se mantienen en el sistema que las creó o, como mucho, se duplican en algún sistema que demandó la realización de dicha observación, o se almacena una parte o un informe de la misma.

Determinados atributos dentro de la información demográfica pueden sufrir varios cambios a lo largo de la vida de una persona. Por ejemplo la dirección del hogar, la del trabajo o el número de teléfono móvil. En un sistema aislado puede considerarse innecesario mantener el valor anterior de estos atributos cuando se produce un cambio, en especial teniendo en cuenta que no suelen ser atributos que se utilicen para la identificación, debido precisamente a su posibilidad de cambio frecuente. No ocurre tanto así con las observaciones clínicas, por ejemplo, donde los diseños, si han sido realizados conforme a los estándares existentes, aseguran un riguroso control de versiones y de auditoría.

La recogida de información demográfica se realiza, en la mayoría de los sistemas, a través de interfaces hombre-máquina y además la mayoría de estos datos son de tipo texto y no pueden codificarse, por lo que es frecuente que los usuarios puedan utilizar distintas cadenas de texto para representar la misma información, o que se cometan

errores tipográficos. Esto indica que incluso desde el momento de su recogida la misma información pueda almacenarse de distintos modos en dos sistemas diferentes.

Las tres características mostradas hasta el momento provocan que las inconsistencias de entidad sean muy probables y que la resolución de los conflictos de entidad en este tipo de información sea especialmente compleja.

Por otro lado, parte de los atributos registrados sobre las personas están ligados a la misma y son fijos o al menos es muy improbable que cambien durante la vida de éstas. Precisamente estos atributos son especialmente útiles para realizar la identificación unívoca de la persona.

La asignación de coeficientes que representen la probabilidad de cambio a los atributos puede ayudar en las tareas de los agentes demográficos, tanto a la hora de resolver los conflictos de entidad como a la hora de la identificación. Precisamente para esta tarea el uso de coeficientes que midan la capacidad que un atributo tiene para identificar a una persona también resulta interesante. Esta capacidad de identificación está ligada a dos factores: la probabilidad de cambio y la probabilidad de que dos personas tengan asignado el mismo valor en ese atributo. Así, los atributos con coeficientes más altos de estabilidad y de identificación son los asignados por alguna administración, como el DNI o el NSS, que de hecho se consideran identificadores, aunque la mayoría de los sistemas suelan gestionar un identificador de persona local.

Para terminar, señalar que determinados atributos de persona pueden tener múltiples valores para la misma persona, por ejemplo se pueden tener dos direcciones postales. Hacer referencia a uno de los valores concretos puede resultar complicado. En ocasiones esta diferenciación puede ser una cuestión semántica, ya que al especificar más la semántica del atributo se puede discriminar a cuál de los valores del mismo se está haciendo referencia. Por ejemplo si una persona tiene dos direcciones, el atributo dirección postal tiene dos valores, pero al referirse a los atributos dirección de trabajo y dirección de casa esa multiplicidad puede desaparecer. Sin embargo existen casos en los que esa diferenciación es imposible, por ejemplo se pueden tener dos números de teléfono móvil y que el sistema no registre ninguna información adicional sobre los mismos; la única forma de distinguir un número del otro es ordenándolos y haciendo referencia al primer o segundo número de teléfono. Esto simplemente muestra que las interfaces de componentes que gestionen este tipo de información tienen que considerar

ambos casos, y que es el tipo de atributo el que condiciona la gestión que se deba hacer del mismo.

Siguiendo la filosofía de modelo dual, propuesta en este trabajo para la elaboración de la vista de información, se va a dividir la vista de información en dos partes. En la primera se detalla el modelo de referencia que podría denominarse la **ontología base o de referencia**. Son conceptos que sirven de contenedores de los más específicos del dominio y que deben dar cabida a cualquier información demográfica y servir de soporte a cualquier componente que maneje información de esta naturaleza. A continuación se detalla la **ontología del dominio** que contiene las clases que representan la semántica concreta de la información demográfica; básicamente se definen los atributos, sus características particulares y los perfiles utilizados en la arquitectura.

Los atributos presentados en este trabajo están basados en los GPICS del CEN, para sistemas que utilicen otra ontología puede ser necesario extender la ontología del dominio con conceptos no considerados, estableciendo las relaciones entre éstos y los previamente existentes. Utilizando los componentes de gestión de ontologías disponibles en la arquitectura se facilita tanto la extensión de la ontología, que aquí se presenta en términos de arquetipos, como que cualquier agente puede realizar las traducciones necesarias para comunicarse con los demás.

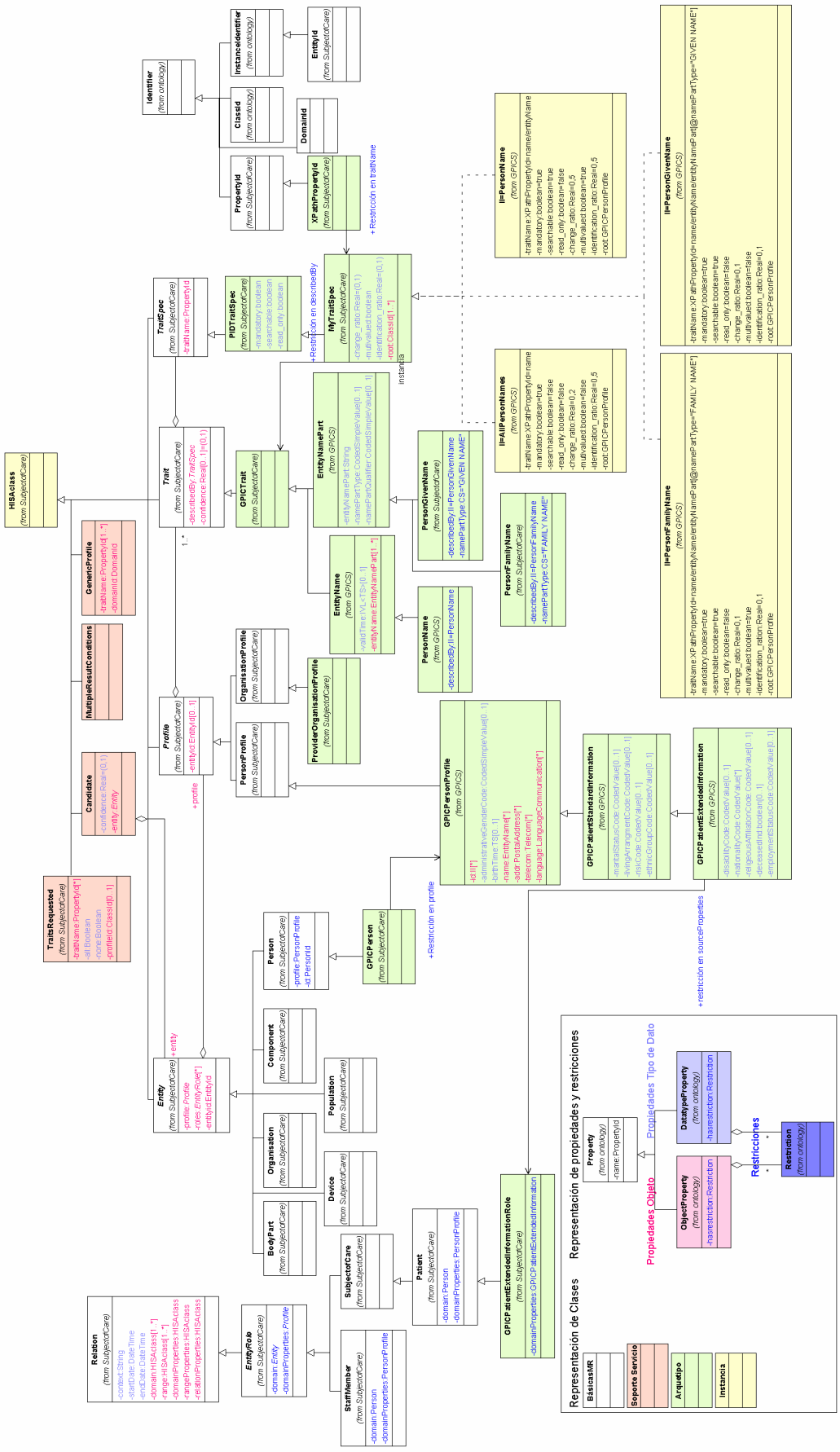


Figura 51. Vista de información. Ontología de referencia y del dominio

7.3.1. El modelo de referencia o la ontología base

El modelo de referencia incluye los conceptos básicos que serán contenedores de la información demográfica y de todos los conceptos asociados a su gestión. Siguiendo la idea de modelo dual serán extendidos en cada dominio u organización. Para definir el modelo se utiliza el modelo de gestión semántica que se propone en esta Tesis, es decir todos los conceptos son clases de la ontología que se relacionan con otras clases o con tipos de datos a través de propiedades. En la figura 51 se muestra tanto este modelo de referencia como el modelo del dominio. Se incluye una leyenda que indica la forma en la que se han representado propiedades, clases y restricciones. Como se puede comprobar algunas clases se utilizan para contener conceptos demográficos y otras conceptos necesarios en la definición de interfaces; considerar estas dos necesidades es una de las ventajas de este modelo con respecto a otros modelos demográficos definidos en el ámbito sanitario.

En cada dominio u organización el modelo de referencia se extenderá para representar la semántica de interés en ese escenario, como se verá en un caso concreto en el siguiente apartado. El modelo de referencia aquí presentado se apoya en el modelo de información del servicio de identificación de personas (PID) de CORBAmed, en la norma 13606, en los GPICS no clínicos y en el modelo de referencia demográfico del openEHR. El primero está más centrado en las necesidades implícitas a la gestión de personas, identificadores e información demográfica, los demás en la representación normalizada de los datos personales. El resultado es fruto de la conciliación de estas referencias y de la realización de las extensiones necesarias para la gestión de atributos con más de un valor y de la información demográfica en entornos federados, que no estaban considerados en el servicio PID.

El modelo de referencia demográfico, como todos los demás, está íntimamente ligado al modelo de información semántico, ya que se basará en los mecanismos de manejo de ontologías para representar los conceptos básicos necesarios para la gestión de datos de personas. Se ha diseñado un modelo muy abierto que permite una reutilización óptima de componentes.

Los conceptos básicos del modelo de referencia son:

- **Entidad (`Entity`):** cualquier actor, sea cual sea su naturaleza, que participe en la organización sanitaria ya sea con un papel activo o pasivo. Una entidad está

caracterizada por un perfil, que es el conjunto de atributos ligado a la misma en cualquier papel que ejecute dentro de la organización, y por un conjunto de relaciones que la vinculan a cualquier otro objeto. Este concepto no aparece en el modelo de información del CEN, pero sí aparece un concepto similar (**PARTY**) en el modelo del openEHR. El término **Entity** en esta Tesis tiene una semántica más abierta que **PARTY**, ya que no se limita a actores con papel activo en la organización.

- **Identificador de Entidad (EntityId):** una entidad en un dominio tiene asignado un identificador que se refiere a la entidad del mundo real a la que representa. Esto es necesario ya que la entidad real tiene que identificarse dentro de cada dominio. Podría pensarse que el identificador de instancia de cualquier clase HISA podría usarse para esta tarea, sin embargo hay que tener en cuenta que dentro de un mismo dominio una entidad real puede estar instanciada varias veces, por lo que el identificador de instancia no proporciona la información necesaria para reconocer que dos objetos son representaciones de la misma entidad del mundo real. Ya que los identificadores no tienen por qué ser universales y distintos dominios de la federación podrían usar distintos identificadores para la misma entidad real es necesario gestionar esto de forma apropiada. Los campos **creationAgent** y **creationUnit** de los atributos de sistema del objeto **Entity** indican quién creó la instancia de entidad, y por tanto pueden servir para indicar el dominio de este identificador de entidad. Otra opción sería incluir un campo dominio en el **EntityId,** que facilita la gestión aunque podría resultar redundante.

- **Atributo (Trait):** es equivalente al concepto **HISAclassAttribute**, representa una propiedad de una clase HISA, pero en este caso se le añade una referencia a un elemento especificación de atributo (**TraitSpec**), que lo describe para facilitar su gestión. Puede incluir una propiedad confianza **(confidence)** que indica la seguridad que se tiene del valor de ese atributo: un uno indicaría que el valor del atributo es seguro. Este campo de confianza se utilizará especialmente en componentes federados que gestionen entidades y sus atributos, como se verá en el modelo de servicio. A un componente que implemente interfaces de gestión de datos demográficos sólo se le podrá

preguntar por un conjunto determinado de atributos que deberán ser identificados por su nombre.

- **Especificación de Atributo (`TraitSpec`):** representa al conjunto de propiedades que caracteriza a un atributo. El modelo de referencia sólo obliga a que exista la propiedad nombre de atributo (**`EntityName`**), dejando a cada implementación que especifique las características de sus atributos como considere necesario. El nombre de atributo (**`EntityName`**) identifica de forma unívoca la semántica del atributo, es decir que se puede considerar un identificador de propiedad. Utilizar los parámetros de especificación de atributos adecuados permite aumentar la lógica de los componentes de gestión de entidades, y en particular de personas, facilitando las tareas de reconciliación de instancias.

- **Perfil (`Profile`):** un perfil es el conjunto de atributos correspondientes a una entidad determinada. Puede incluir también la identificación de la entidad a la que se refiere, aunque no es obligatorio. Se pueden crear arquetipos de perfiles, de manera que si se indica que un perfil debe seguir un arquetipo determinado éste debe cumplir las especificaciones impuestas por el arquetipo: propiedades y restricciones. En este caso el identificador de arquetipo es el identificador de la clase de perfil, que no se debe confundir con el identificador de la instancia del perfil.

- **Relación (`Relation`):** una relación vincula varios objetos de la organización especificando una semántica determinada a ese enlace. Es un concepto equivalente al de propiedad de una clase de una ontología, y al igual que en ese caso se distingue entre el dominio y el rango. Dentro de una relación se pueden indicar propiedades específicas de los objetos relacionados que se restringen al ámbito de dicho vínculo. Un tipo particular de relación es el rol de Entidad (**`EntityRole`**), que indica la relación de una entidad con cualquier otro objeto de la arquitectura, y donde se incluye un perfil específico para la entidad cuando ésta asume este papel en concreto. Haciendo la analogía con el openEHR, sería equivalente al concepto **`PartyRelationship`**, donde dominio y rango serían equivalentes a **`source`** y **`target`**.

- **Relación Sujeto de Cuidado (`SubjectofCare`):** un tipo particular de rol de entidad es el de sujeto de cuidado. Hay que tener en cuenta que un sujeto de cuidado puede ser cualquier entidad susceptible de ser atendida dentro de la organización y que este concepto está abierto a muchas posibilidades; personas, órganos, poblaciones, animales, etc.

- **Persona (`Person`):** es una especialización de Entidad, lo que indica que los procedimientos de gestión de perfiles y relaciones, genéricos para cualquier entidad, son aplicables a una persona. Esta es una de las grandes ventajas de este modelo: que facilita la reutilización de componentes. En este caso el identificador de entidad es un identificador de persona, el equivalente al **`PersonId`** del PID de CORBAmed.

- **Relación Paciente (`Patient`):** es un tipo particular de relación Sujeto de Cuidado en la que la entidad a la que se aplica es una persona. El perfil específico de la persona como paciente se incluye en esta relación.

- **Relación miembro de plantilla (`Staffmember`):** tipo especial de rol de entidad, aplicable sólo a personas, que indica que esta persona es un trabajador de la organización. Los atributos característicos de la persona como personal de la organización van incluidos en esta relación.

También es necesario incluir algunos conceptos que servirán de apoyo en la definición de las interfaces de los componentes de gestión de personas:

- **Perfil Genérico (`GenericProfile`):** conjunto de nombres de atributos que forman una clave única en un dominio determinado.

- **Atributos pedidos (`TraitsRequested`):** conjunto de nombres de atributos de la entidad solicitados en la interacción entre dos componentes de la arquitectura.

- **Candidato (`Candidate`):** es una entidad a la que se le añade un campo **confianza (`confidence`)** cuyo valor está entre cero y uno. Se devuelve como resultado de una búsqueda de entidad y la confianza indica la precisión con la que la entidad devuelta cumple los criterios de búsqueda. Una confianza de uno indica que la entidad devuelta es seguro la buscada.

- **Condiciones para resultado múltiple (`MultipleResultConditions`):** cuando el resultado de un método puede ser más de un elemento, permite indicar cómo se quieren gestionar estos elementos. Es un requisito especificado en HISA para las interfaces en las que manejan conjuntos de elementos.

7.3.2. La ontología del dominio

Los componentes que permiten la gestión de instancias conforme al modelo de referencia presentado serán fundamentales en las actividades relacionadas con la gestión de personas y otras entidades. Sin embargo la semántica y características concretas de los atributos, como se ha podido comprobar, no están especificadas en el modelo de referencia. Para aportar la semántica que falta es necesario extender la ontología federada con conceptos más específicos del dominio. En el modelo dual del openEHR esto sería lo que se conoce como arquetipos. La figura 51 presenta en color verde estos arquetipos y en amarillo instancias concretas de los mismos.

La ontología del dominio propuesta en esta Tesis Doctoral comienza especificando una semántica de atributos basada en los GPICS del CEN, que se pueden ver en la figura 16, principalmente porque al enmarcarse en el entorno europeo el CEN es el mejor referente y porque las principales normas de informática sanitaria, incluida la 13606 donde se define el modelo de historia clínica, hacen referencia a los GPICS no clínicos cuando se trata de la normalización de la información demográfica. Todos los atributos presentados en la figura 51 son especializaciones del concepto **`Trait`** del modelo de información, donde sólo se muestra una pequeña parte de los mismos para simplificar el gráfico. El mecanismo es totalmente abierto, de manera que pueden especificarse otros atributos distintos a los de GPICS y establecerse equivalencias y otras relaciones entre conceptos, gracias a las actividades de semántica que permiten una gestión eficiente de la extensión de la ontología en cada dominio u organización.

Por otro lado se definen arquetipos de perfiles que se corresponden con los especificados en los GPICS para las clases **`Person`**, **`PatientStandardInformation`**, **`PatientExtendedInformation`** y **`HealthcareProfesionalRole`**, que se mostraron en la figura 16 del capítulo correspondiente a los servidores demográficos. Estos arquetipos son simplemente una especialización del concepto **`Profile`**, donde se establecen restricciones sobre las propiedades que se pueden encontrar en esta clase. El perfil de persona estará asociado a

la clase **Person,** y los específicos de cada rol a la relación que los representa. Del mismo modo podrían definirse perfiles personalizados siguiendo el mismo mecanismo. El identificador de la clase perfil adecuada juega el papel de identificador de arquetipo.

Cada atributo de un perfil será una especialización de **Trait**, en este caso basados en los atributos de los GPICS no clínicos. Otra especialización que se realiza es relativa al nombrado de atributos. El nombre de un atributo es una especialización de un nombre de propiedad, y en este caso se restringe que el sistema de nombrado debe seguir el formato XPath. Con este sistema de nombrado se puede identificar un atributo cualquiera de la clase persona, sea cual sea el nivel al que se encuentre, y se pueden distinguir también los atributos multievaluados. Precisamente XPath permite la especificación concreta de un valor de un atributo multievaluado a partir de alguna de sus propiedades. Uno de los ejemplos mostrados en la figura es la especificación del atributo **PersonFamilyName**. Dado que el nombre de una persona (de la clase **EntityName**) puede tener varios **entityNamePart** (de la clase **EntityNamePart**), se puede identificar el atributo apellido como el **entityNamePart** que tenga dentro de la propiedad **namePartType** el valor **FAMILYNAME**. Así el nombre de este atributo se podría especificar en XPath como:

name/entityName/entityNamePart[@namePartType=”FAMILYNAME”]

Teniendo en cuenta que el nombre de una persona es multievaluado la identificación del apellido dentro de un nombre concreto de la persona sería:

PersonId/name[EntityNameId]/entityName/entityNamePart[@namePartType=”FAMILYNAME”]

El uso de códigos para identificar el **namePartType** permite introducir fácilmente nuevos tipos, como por ejemplo el segundo apellido, en caso de que fuera necesario. Los GPICS del CEN utilizan continuamente este recurso, por lo que la mayoría de los atributos demográficos que podamos encontrar en un dominio y que no estén considerados en los GPICS no clínicos pueden especificarse con este mecanismo de extensión de códigos. Dentro de la arquitectura, los componentes de gestión de códigos permitirán el manejo e interpretación de los mismos.

Otra especialización interesante, que se presenta en la figura 52, es la de la clase **TraitSpec**. Esta especialización está basada en el PID de CORBAmed y en algunas necesidades identificadas para la federación de sistemas de gestión demográficos. Se

hace en dos niveles, **`PIDTraitSpec`** y **`MyTraitSpec`**, para identificar claramente qué atributos se heredan de la especificación PID y cuales son específicos para la gestión de federaciones entre componentes de gestión demográfica, que son los incluidos en esta Tesis. Estos son:

- **`change_ratio`:** valor entre 0 y 1 que indica la probabilidad de cambio en el valor del atributo. Por ejemplo la probabilidad de cambio del número de teléfono móvil es mucho mayor que la de cambio del primer apellido.
- **`multivalued`:** si es cierto indica que este atributo es un conjunto, es decir que puede tener múltiples valores. Por ejemplo las direcciones postales de una persona pueden ser varias, mientras que su fecha de nacimiento sólo tiene un valor.
- **`identification_ratio`:** valor entre 0 y 1 que indica la probabilidad de que el valor de este atributo identifique unívocamente una entidad. Por ejemplo, el DNI tendrá un identification_ratio muy próximo a 1, mientras que el sexo tendrá un valor cercano a 0.

También se añade la restricción de que el nombre del atributo esté en formato XPath, como se había indicado antes, y se incluye un atributo para identificar en qué clases puede aparecer este atributo. Esta exhaustiva especialización de **`TraitSpec`** permite realizar una gestión muy eficiente de entidades en un entorno federado y facilita las tareas de reconciliación de instancias.

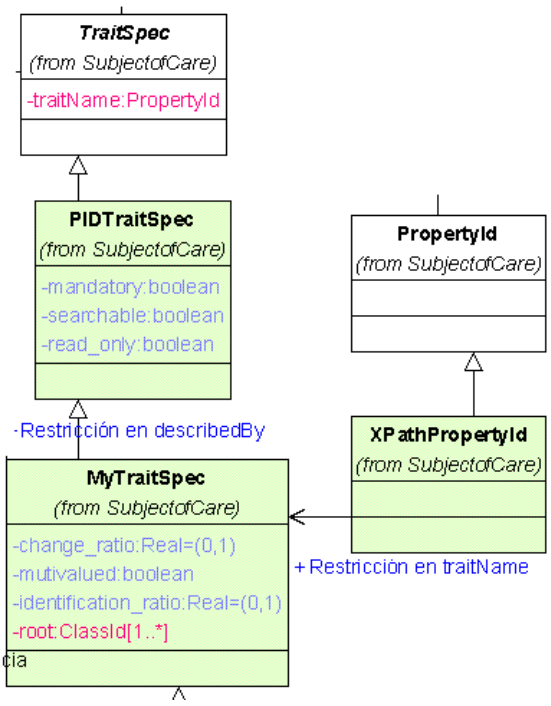


Figura 52. Ejemplo de arquetipo de TraitSpec

La figura 51 incluye el modelo de referencia presentado más la especialización de algunos conceptos como arquetipos, a modo de ejemplo.

En HISA no sería necesario incluir este nivel del modelo de información, cuyo desarrollo quedaría en manos de cada organización, aunque en esta Tesis se presenta como ejemplo de una posible materialización de este nivel de descripción de ontología.

7.4. Vista computacional

La vista computacional debe ser una extensión de la de HISA y, en este caso, está basada en las interfaces definidas en el PID de CORBAmed. Para la especificación de los componentes, interfaces y servicios debería utilizarse el lenguaje computacional normalizado por HISA, si esta normalización tiene lugar. En este caso se utiliza la ontología de descripción propuesta en esta Tesis Doctoral. Debido a que en este lenguaje se incluyen todos los conceptos de ODP, para hacer la descripción computacional más sencilla no se van a utilizar todos los recursos del lenguaje propuesto. Así, por ejemplo, quedan fuera del alcance de este trabajo la definición de clases para representar aspectos como el contrato de entorno o el comportamiento. El modelo semántico permite que cada arquitectura defina los conceptos de interés que no estén especificados en HISA, facilitando por tanto el mecanismo de extensión.

Se van a especificar los siguientes conceptos normalizados en ODP:

- **`Interfaz`:** una interfaz representa una parte del comportamiento del componente relacionado con un subconjunto particular de sus interacciones posibles. Cada interfaz está identificada por un conjunto de interacciones en las que el objeto puede participar. Será necesario especificar las firmas de interfaz, el contrato de entorno y el comportamiento de un objeto para que éste esté completamente especificado. En este trabajo se profundiza sólo en las firmas de interfaz que facilitan la interacción con el objeto.

Tabla I. Clasificación de interfaces

Según la información gestionada		**Según el dominio**	
Básicas	gestión elemental de objetos del modelo	**Autónomas**	actúan sobre un único dominio de la federación
Complejas	permiten gestión más avanzada	**Federadas**	actúan sobre varios dominios de la federación

- **`Servicio`:** [126] un servicio es un conjunto de capacidades proporcionadas por un objeto en una interfaz computacional. La descripción del servicio tal y como está planteada en el lenguaje de especificación desarrollado en esta Tesis Doctoral facilitará la selección de las interfaces adecuadas para lograr los objetivos buscados.

A continuación se detallan las interfaces definidas que se han clasificado según muestra la tabla i.

7.4.1. Interfaces básicas y autónomas

Son componentes que actuarán sobre un único dominio y sólo tienen tareas de gestión elemental de objetos del modelo de información.

Dado que una persona es un tipo especial de entidad, para definir los componentes e interfaces de gestión de personas se puede comenzar con los de gestión de entidades y sus atributos, y en particular con la definición de los métodos básicos. Los componentes que gestionan personas serán una especialización de éstos.

Las interfaces básicas y autónomas implicadas en las tareas de gestión demográficas se han considerado sólo del tipo operación. Primero se han definido los parámetros, que se muestran en la tabla ii, y a continuación las terminaciones, representadas en la tabla iii.

Finalmente la tabla iv incluye las firmas de invocación e interrogación definidas. Se ha seleccionado sólo un pequeño número de operaciones, las más importantes, pero utilizando el lenguaje basado en ODP que se ha definido y algún editor como Protègè, tanto la extensión como la edición de firmas resultan muy sencillas.

Tabla **II**. Parámetros definidos para la gestión demográfica

Firma de parámetro	**Nombre**	**Tipo de Dato**
01	Requested Profile	ProfileId
02	Requested Entity	EntityId
03	Requested Person	PersonId
04	Known Profile	Profile
05	Requested Role	RoleId
06	Queried Entity	Entity
06_SET	Queried Entities	Entity_set
07	Queried Person	PersonId
08	Queried Role	EntityRole
09	Detailed Profile	Profile
10	Generic Profile	GenericProfile
10_SET	Generic Profiles	GenericProfile_set
11	Unknown ProfileId exception	Exception
12	Unknown Profile exception	Exception
13	Queried EntityId	EntityId
13_SET	Queried EntityIds	EntityId_set
14	Requested Traits	TraitRequested
15	Queried RoleId	RoleId
15_SET	Queried RoleIds	RoleId_set
16	Unknown RoleId Exception	Exception
17	New Entity	Entity

Firma de parámetro	Nombre	Tipo de Dato
18	Entity Exists exception	Exception
19	Finalised Operation	Confirm
20	Uncompleted Profile Exception	Exception
21	Modified Trait	II
22	New Trait	Trait
23	Unknown Trait	Exception
24	Queried TraitId	TraitId
24_SET	Queried TraitIds	TraitId_set
25	Unknown entity Id	Exception
26	Requested Trait	TraitId
27	Queried Trait	Trait
27_SET	Queried Traits	Trait_set
28	Domain Id	DomainId
29	UnknownDomain Exception	Exception
30	New entity Role	EntityRole
31	Duplicated Entity Role exception	Exception
32	Duplicated Trait	Exception
33	Type of Entity	ClassId

Tabla III. Terminaciones

Firma terminación	Nombre	Firma parámetro
01	LIST ENTITIES TERMINATION	06_SET
02	LIST ENTITY IDS TERMINATION	13_SET
03	DETAIL ENTITY TERMINATION	06
04	LIST ENTITY ROLE IDS TERMINATION	15_SET
05	DETAIL ENTITY ROLE TERMINATION	08
06	ADD ENTITY TERMINATION	19

Firma terminación	Nombre	Firma parámetro
07	UPDATE ENTITY TERMINATION	19
08	UPDATE PROFILE TERMINATION	19
09	UPDATE TRAIT TERMINATION	19
10	ADD TRAIT TERMINATION	19
11	LIST TRAIT IDS TERMINATION	24_SET
12	LIST TRAIT TERMINATION	27_SET
13	LIST GENERIC PROFILES TERMINATION	10_SET
14	ADD GENERIC PROFILE TERMINATION	19
15	ADD ENTITY ROLE TERMINATION	19
EXCEPTION_11	UNKNOWN PROFILE ID TERMINATION	11
EXCEPTION_12	UNKNOWN PROFILE TERMINATION	12
EXCEPTION_16	UNKNOWN ROLEID TERMINATION	16
EXCEPTION_18	ENTITY EXISTS TERMINATION	18
EXCEPTION_20	UNCOMPLETED PROFILE TERMINATION	20
EXCEPTION_23	UNKNOWN TRAIT TERMINATION	23
EXCEPTION_25	UNKNOWN ENTITY ID TERMINATION	25
EXCEPTION_29	UNKNOWN DOMAIN TERMINATION	29
EXCEPTION_31	DUPLICATED ENTITY ROLE TERMINATION	31
EXCEPTION_32	DUPLICATED TRAIT TERMINATION	32

Tabla **IV**. Operaciones (tipo interrogación) e invocaciones

Firma Interro-gación	Firma Invocación		Firma Terminación	Tipo
	Nombre	Parámetro		
01_A	LIST ENTITIES INVOCATION	01,04	01,EXCEPTION_11,EXCEPTION_12	list
01_B	LIST ENTITIES INVOCATION	04,14	01,EXCEPTION_12,EXCEPTION_23	list
01_C	LIST ENTITIES INVOCATION	01,04,33	01,EXCEPTION_11,EXCEPTION12	list
01_D	LIST ENTITIES INVOCATION	04,14,33	01,EXCEPTION_12,EXCEPTION_23	list

	Firma Invocación			
02_A	LIST ENTITY IDS INVOCATION	04	02,EXCEPTION_12	list
02_B	LIST ENTITY IDS INVOCATION	04,33	02,EXCEPTION_12	list
03_A	DETAIL ENTITY INVOCATION	01,02	03,EXCEPTION_11,EXCEPTION_25	detail
03_B	DETAIL ENTITY INVOCATION	02,14	03,EXCEPTION_23,EXCEPTION_25	detail
04	LIST ENTITY ROLE IDS INVOCATION	02	04,EXCEPTION_25	list
05	DETAIL ENTITY ROLE INVOCATION	05	05,EXCEPTION_16	detail
06	ADD ENTITY INVOCATION	17	06,EXCEPTION_18,EXCEPTION_20	add
07	UPDATE ENTITY INVOCATION	02,17	07,EXCEPTION_20	update
08	UPDATE PROFILE INVOCATION	02,04	08, EXCEPTION_20	update
09	UPDATE TRAIT INVOCATION	21,22	09,EXCEPTION_23	update
10	ADD TRAIT INVOCATION	02,22	10,EXCEPTION_23,EXCEPTION_32	add
11	LIST TRAIT IDS INVOCATION	02,26	11,EXCEPTION_25	list
12	LIST TRAIT IDS INVOCATION	02	12,EXCEPTION_23,EXCEPTION_25	list
13_A	LIST GENERIC PROFILES INVOCATION	33	13	list
13_B	LIST GENERIC PROFILES INVOCATION	28,33	13	list
14_A	ADD GENERIC PROFILE INVOCATION	10,28,33	14,EXCEPTION_23,EXCEPTION_29	add
14_B	ADD GENERIC PROFILE INVOCATION	10,33	14,EXCEPTION_23,EXCEPTION_29	add
15	ADD ENTITY ROLE INVOCATION	30	15,EXCEPTION_12,EXCEPTION_23, EXCEPTION_31	add

Una vez especificadas las firmas de operación se pueden definir múltiples firmas de interfaz operativa en función de la combinación de operaciones que se elija, como se muestra en la figura 53, donde se presentan dos interfaces con causalidad complementaria, y la figura 54, donde todas las interfaces tienen causalidad servidor.

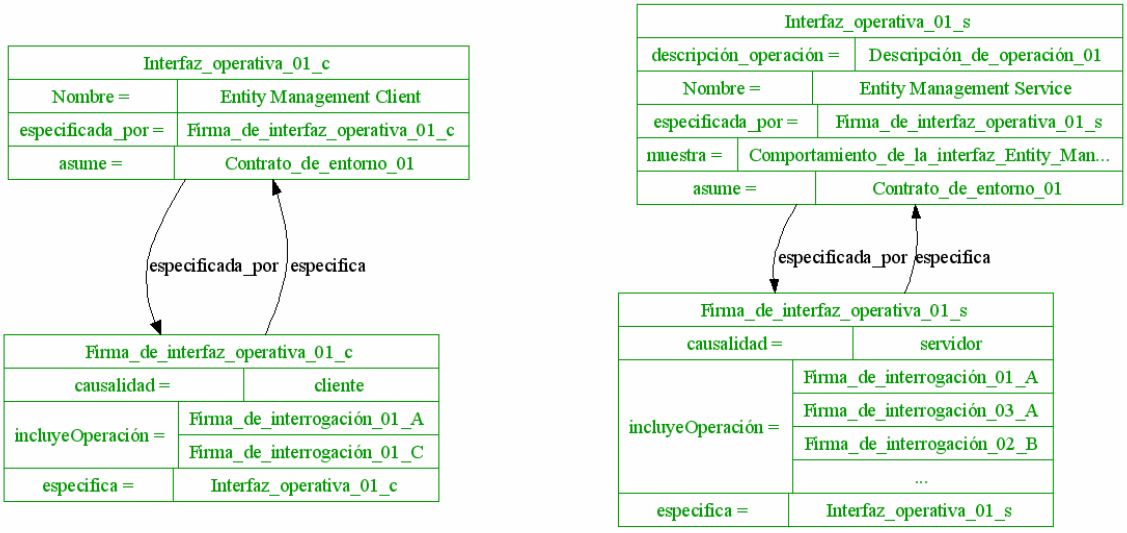


Figura 53. Ejemplo de interfaces con causalidad complementaria.

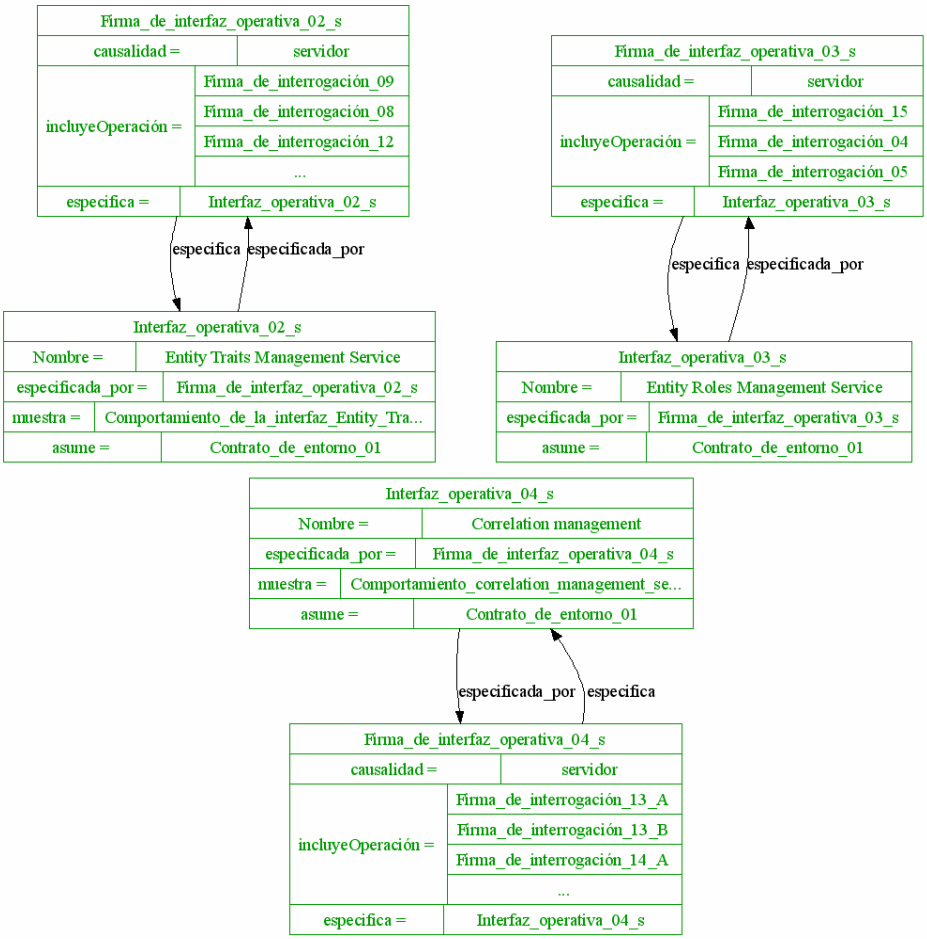


Figura 54. Otras interfaces de ejemplo, todas con causalidad servidor

Una vez definidas las interfaces se pueden definir servicios, como a modo de ejemplo se muestra en la figura 55, donde se describen servicios para la gestión de entidades y para la gestión de atributos utilizando algunas de las interfaces presentadas anteriormente. Puede comprobarse que en ambas definiciones el dominio que se gestiona es el local.

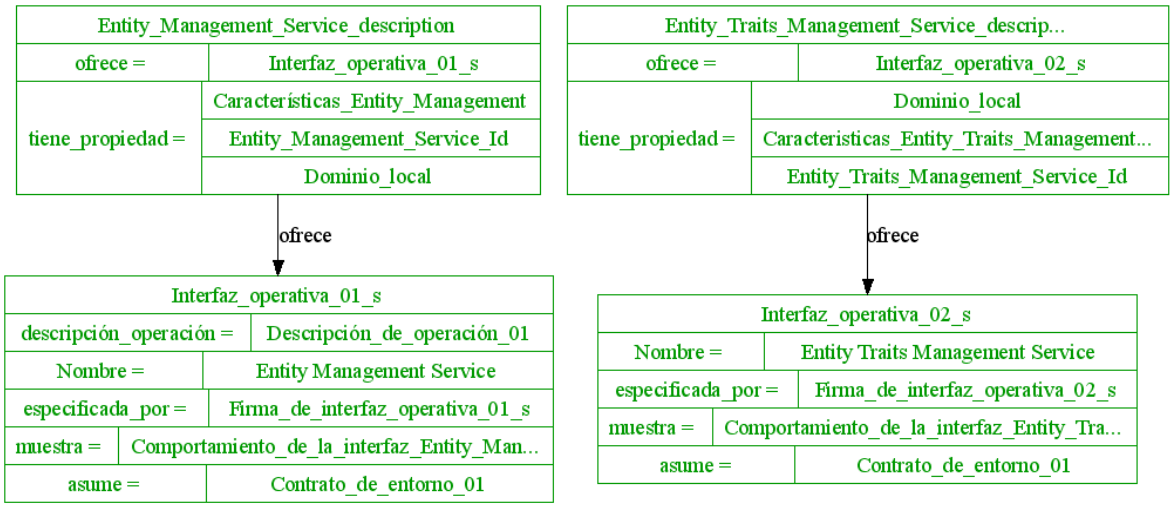


Figura 55. Definición de servicios de gestión de entidad y de gestión de atributos

Como ejemplo más detallado se presenta la descripción del servicio para gestión de entidades, mostrada en la figura 56. Las propiedades que se han definido para describir el servicio son el dominio sobre el que actúa, un identificador único y un tercero denominado **`características`** donde se incluye: en qué actividades de la organización se usa, de qué tipo son los métodos implementados y qué tipo de objetos gestiona. Estas propiedades servirán para seleccionar adecuadamente los servicios a utilizar, y por supuesto podrían extenderse para proporcionar una selección más precisa.

Se puede comprobar que como el servicio de ejemplo es autónomo sólo actúa sobre un dominio (**`Id_domino_1`**). Dentro de éste servicio se detalla la operación para listar entidades, mostrada en la figura 57.

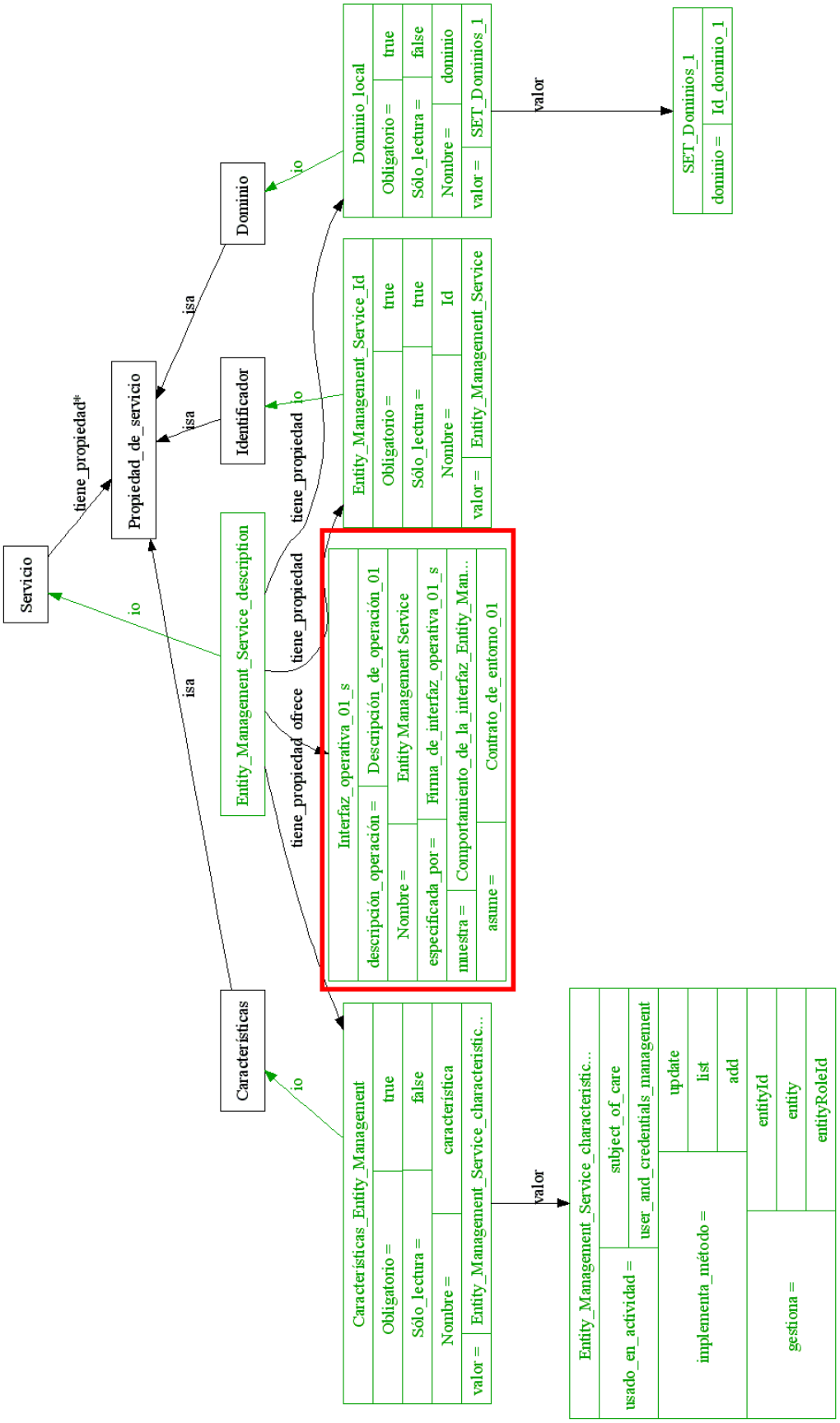


Figura 56. Descripción del Servicio para gestión de entidades

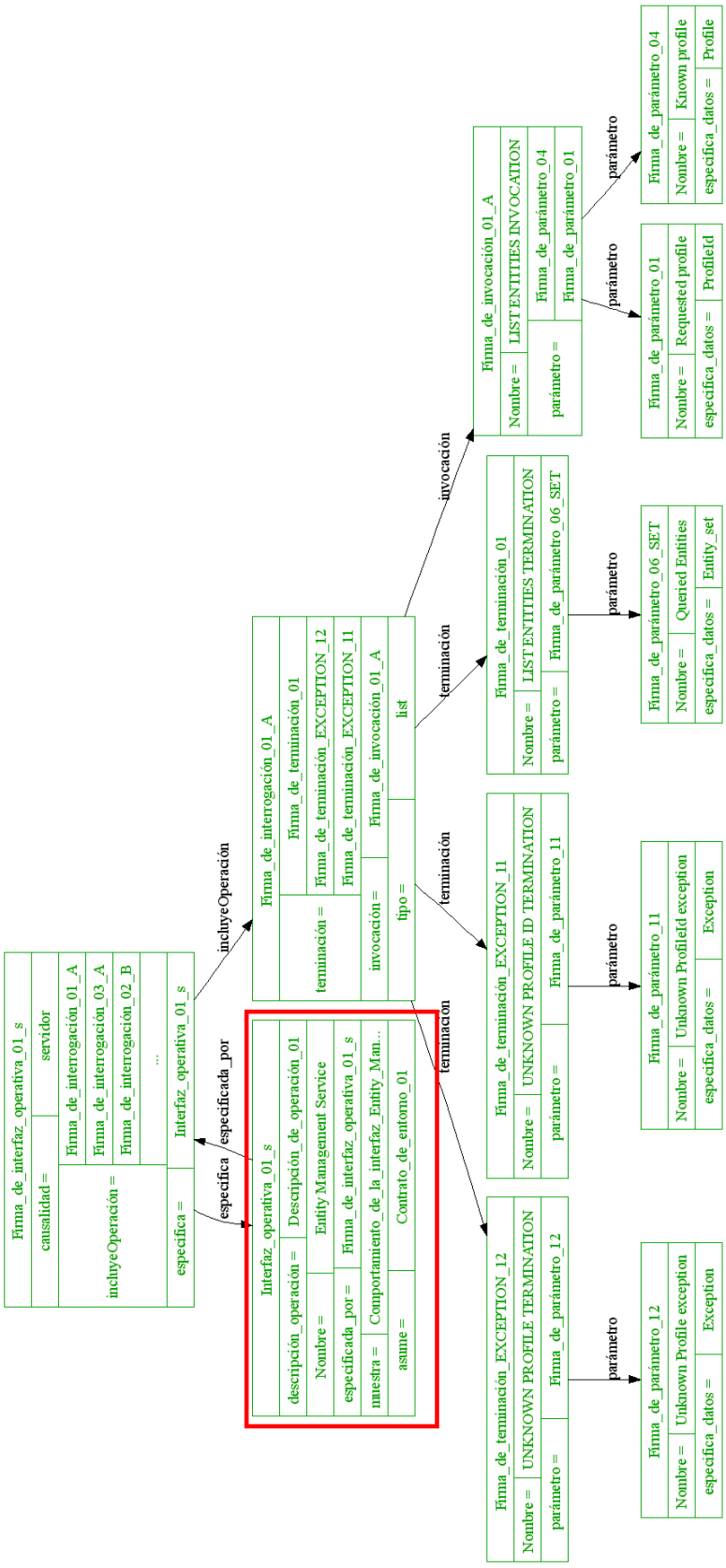


Figura 57. Resumen de la especificación de la operación para listar entidades (pertenece a la interfaz del servicio de gestión de entidades)

Como puede comprobarse, las únicas operaciones en la que interviene el identificador de dominio (parámetro 28) son las que se utilizan para la gestión del perfil genérico (**`LIST GENERIC PROFILES INVOCATION`** y **`ADD GENERIC PROFILE INVOCATION`**). En todas las demás no se especifica el dominio por lo que se entiende que no existe ambigüedad en los extremos de las vinculaciones respecto al dominio de interés. En ninguna de las operaciones básicas diseñadas se manejan confianzas de candidatos: sólo se devuelven valores que cumplen exactamente con la búsqueda.

7.4.2. Interfaces complejas y autónomas

Estas funciones permiten tareas más complejas que la gestión directa de un objeto de la arquitectura, pero siempre actuando únicamente dentro de un dominio. Estarían dentro del grupo de actividades del sujeto de cuidado tanto las actividades de gestión de pacientes como las de gestión de contactos. En esta Tesis se tratan las actividades de gestión de personas en general, y no únicamente pacientes, que se consideran parte del flujo del sujeto de cuidado pero que podrían ser utilizadas en otros grupos de actividades.

Las principales interfaces de gestión de personas pueden extraerse del servicio PID de CORBAmed y de las descritas en la primera versión de la vista computacional de HISA. Algunas de las tareas especificadas en ambos estándares son prácticamente iguales que las mostradas como básicas y autónomas en el apartado anterior y otras son simplemente una combinación de varias de éstas. La descripción mostrada en el comportamiento debe incluir la reutilización que se hace de interfaces básicas, lo que puede ayudar a decidir si se establecen o no vínculos con esa interfaz.

- **`Add Person in Role`:** permite añadir una persona nueva al sistema pero indicando el rol que la persona juega en el mismo. Pueden existir distintas versiones de esta interfaz como **`registrar paciente`**, **`registrar jefe de servicio`** o **`registrar usuario`**. Un objeto conseguirá ofrecer estas interfaces cuando establezca vínculos en interfaces de otros objetos con tareas como las de añadir entidad o añadir rol de entidad. Por supuesto será tarea del objeto comprobar que la persona no estaba ya en el sistema y que se han incluido todos los atributos necesarios en los perfiles de rol y de entidad introducidos.

- **`Merge Persons:`** permite fusionar dos objetos persona en caso de identificar que se corresponden con la misma persona física. Esta fusión implica una reconciliación de las dos instancias, es decir resolver los posibles conflictos de atributos, utilizar un único identificador de entidad, etc. Será necesario utilizar interfaces ofrecidas por otros objetos que realicen tareas como actualizar entidades, roles de entidad o perfiles.
- **`List PersonID in Role:`** permite localizar una persona, y obtener su identificador de entidad, pero utilizando para la búsqueda datos conocidos de un rol determinado. Si la persona es localizada en el sistema pero no tiene ese rol no se devolverá como un resultado válido. Pueden existir distintas especializaciones como **`localizar paciente`**, **`localizar jefe de servicio`** o **`localizar usuario.`** Un objeto conseguirá ofrecer esta interfaz cuando establezca vinculaciones con otros objetos para utilizar interfaces ofrecidas por éstos; como las de listar identificadores de entidad o listar entidades.
- **`Detail Person in Role:`** permite obtener los datos personales de una persona pero para un rol específico. Pueden existir distintas especializaciones como **`consultar paciente`**, **`consultar jefe de servicio`** o **`consultar usuario.`** Para ofrecer esta interfaz es necesario establecer vínculos con objetos que ofrezcan interfaces como las de detallar entidad o detallar rol de entidad.

7.4.3. Interfaces federadas

Las funciones federadas que se presentan aquí son sólo de consulta. Básicamente la funcionalidad debe ser la misma que en el caso de operaciones autónomas, aunque en esta ocasión se utilizará más de un dominio, por lo que será necesario realizar tareas de reconciliación de entidades y valores de atributos. De manera que los objetos que ofrezcan estas interfaces se vincularán a objetos que implementen las operaciones descritas en los apartados anteriores y realizarán un esfuerzo extra para reconciliar entidades y gestionar identificadores de entidad que se mantendrán, al menos, durante una sesión. Se podrán distinguir interfaces federadas básicas, que gestionan directamente un objeto del modelo de información, aunque en este caso el objeto se ha creado a partir de información procedente de varios dominios, e interfaces complejas que realizan tareas de gestión más avanzadas.

Dado que las funcionalidades básicas de estas interfaces son prácticamente las mismas que las ya mostradas y que la existencia de varios dominios es totalmente transparente a los usuarios de estas interfaces federadas, los métodos serán muy similares.

Las tareas principales de estas interfaces serán:

- La localización de una persona determinada en cada uno de los sistemas y la obtención de un identificador de entidad, en este caso de persona, que servirá para identificar a ésta en un conjunto de dominios durante, al menos, una sesión.
- La obtención de la entidad, de tipo persona, con el perfil solicitado y resolviendo los conflictos de valores de atributos que pudieran surgir con la menor intervención del usuario.

Las tareas de registro federadas quedan fuera del alcance este trabajo ya que se considera que el registro, en la mayoría de las federaciones, será una tarea local.

Se utilizarán en estas interfaces los conceptos de confianza de entidad y de atributos. El cálculo de esas confianzas estará íntimamente ligado a la especificación de **`TraitSpec`** en la arquitectura. La forma de realizar este cálculo debe poder ser consultado por los elementos de la arquitectura que vayan a hacer uso de esos parámetros de confianza. A través del comportamiento del servicio se puede especificar formalmente el uso que se realiza de estos parámetros. Para manejar el concepto de confianza en la entidad recuperada se ha introducido el concepto de **`candidato`**. De manera que en este caso en lugar de utilizar el método **`ListEntities`**, que devolvería un conjunto de entidades, se utilizará **`ListCandidates`**, que devolverá un conjunto de candidatos.

La búsqueda estará guiada desde el primer momento por los atributos que forman parte del perfil genérico de la federación, es decir que se utilizan estos atributos para realizar la búsqueda de la persona dentro de la federación. Las principales operaciones de la interfaz de identificación de personas serán:

- **`Detail Federated Generic Profile`:** que proporciona el perfil genérico correspondiente a los dominios sobre los que se quiere hacer la búsqueda. Este perfil genérico federado es la unión de los atributos de todos los perfiles genéricos de una serie de dominios. Los dominios serán aquellos sobre los que se quiere hacer la búsqueda de la persona, y los identificadores de los

dominios pueden pasarse como parámetros o bien dependerá de los credenciales del agente invocador y del usuario en nombre del cual actúe éste. El objeto que proporcione este perfil genérico federado debe interactuar con otros que proporcionen los perfiles genéricos de los dominios implicados en la búsqueda, a través del método **`List Generic Profiles`**.

- **`List Candidates`:** devuelve los candidatos que cumplan el perfil introducido. Cuantos más atributos del perfil federado genérico se utilicen en el perfil de búsqueda, más eficiente será la misma. El candidato incluye un perfil, y los valores de los atributos de este perfil deben ser verificados por el componente que está realizando la búsqueda. Los atributos de este perfil tienen un indicador de confianza que estará entre 0 y 1, indicando el 1 que el valor del atributo es seguro el señalado. El candidato incluye otro indicador de confianza que será más próximo a 1 cuanto más exacta haya sido su identificación. En principio el de mayor confianza sería el único de interés para el cliente, aunque los mecanismos utilizados para el cálculo de los indicadores de confianza en los atributos y en los candidatos deben estar formalizados en la descripción de comportamiento de la interfaz, con la finalidad de facilitar al componente que utilice esta interfaz la comparación de candidatos para seleccionar el correcto y verificar los atributos.
- **`Verify Traits`:** se envía el candidato seleccionado pero en esta ocasión sólo se incluyen en el perfil los atributos cuyo valor es verificado por el usuario. De manera que si éstos tenían una confianza menor que uno a partir de ese momento pasan a tener confianza 1, mientras que los atributos no verificados no modifican su confianza.
- **`Detail Person`:** devuelve una entidad de tipo persona con el perfil o el conjunto de atributos solicitado en la invocación, que incluye también el identificador de entidad del candidato seleccionado en el paso anterior.
- **`Detail Person in Role`:** devuelve los datos específicos de una persona para un papel determinado en la federación. Pueden existir especializaciones como **`Detallar Paciente`** o **`Detallar Usuario`**

Esta interfaz estará proporcionada por un agente de la arquitectura que se podría denominar **`PMI`** (**`Person Master Index`**) y que tiene toda la "inteligencia"

necesaria para realizar la reconciliación de entidades y valores de atributos. Este componente establecería vínculos con objetos que implementaran las interfaces mostradas en los apartados anteriores en cada uno de los dominios implicados en una búsqueda.

7.4.4. Comportamiento

Antes de pasar a la exposición detallada del comportamiento para la identificación de entidades en una federación es necesario aclarar un par de conceptos que se van a utilizar para realizar esta descripción.

Mecanismos de identificación de entidad:

Se distinguen tres variantes al realizar la identificación de una entidad en un dominio de la federación:

- **`Hard Matching`:** se denomina **`hard matching`** a la localización de una entidad en un dominio utilizando como perfil de búsqueda alguno de los perfiles genéricos de este dominio y teniendo todos los atributos del perfil de búsqueda confianza 1. Siempre se devolverá un único candidato, si es que existe, y la confianza asignada al mismo será 1.
- **`Provisional Hard Matching`:** se denomina **`provisional hard matching`** a la localización de una entidad en un dominio utilizando como perfil de búsqueda alguno de los perfiles genéricos de este dominio pero en este caso alguno de los atributos pasados para la búsqueda tiene confianza menor que 1, generalmente porque no han sido introducidos directamente por el usuario que realiza la búsqueda. Siempre se devolverá un único candidato, si es que existe, y la confianza asignada al mismo se considerará menor que 1.
- **`Soft Matching`:** se denomina **`soft matching`** a la localización de una entidad en un dominio utilizando como perfil de búsqueda una combinación de atributos que no coincide con ningún perfil genérico del dominio. En este caso el número de candidatos devuelto puede ser mayor que uno, y la confianza que se asigna a los mismos decrecerá.

Resolución de conflictos de atributos:

A la hora de resolver los conflictos de determinadas entidades y valores de atributos obtenidos de distintos dominios, pero correspondientes a la misma entidad del mundo real, será necesario considerar, entre otros, los siguientes parámetros de atributo:

- **La cardinalidad del atributo:** si el atributo puede tener más de un valor es posible que dos valores, aunque no coincidan, sean aceptables y de interés.
- **Mecanismo de comparación del atributo:** la comparación de valores de atributos será distinta en función del tipo. Por ejemplo, comparar dos cadenas de texto largas es distinto a comparar dos fechas. Deben desarrollar algoritmos que optimicen estas comparaciones para cada tipo de atributo en particular.
- **Discriminación de atributos:** a la hora del descarte de valores de atributos, especialmente cuando el tipo de los mismos es complejo, puede ser interesante incluir en los mismos datos selectivos como el periodo de validez, que puede servir para discriminar el valor de uno de ellos a favor de otro. Por ejemplo, si el momento actual no pertenece al periodo de validez de la dirección postal obtenida del dominio 1, se puede descartar ésta para quedarse sólo con la del dominio 2, en la que la fecha actual entra dentro del periodo de validez.
- **La probabilidad de cambio del atributo:** atributos cuyo valor es muy estable en el tiempo, por ejemplo el sexo, son especialmente apropiados para el descarte de alguno de los candidatos seleccionados y para aumentar la capacidad de identificación.
- **La probabilidad de duplicidad del valor del atributo:** atributos cuyo valor está muy poco duplicado, como el DNI, son muy apropiados para la identificación de entidades o el descarte de las mismas.

Todos estos parámetros que caracterizan al atributo, y todos aquellos que se consideren necesarios, deberían estar detallados en la clase **`TraitSpec`** asociada al atributo y que lo describe en detalle, como se mostraba en la figura 52.

Una vez definidos estos conceptos pasamos a describir el flujo de trabajo que se establece entre los distintos componentes de la arquitectura. Para detallar el comportamiento se ha elegido utilizar un diagrama de secuencia UML. Sin embargo, para que un agente software pudiera entender este comportamiento y tomar decisiones de selección, vinculación, etc, éste debería ser detallado en un objeto del tipo

`Comportamiento`, incluido en la ontología ODP. Existen varios lenguajes que permiten formalizar el comportamiento, por ejemplo OWL-S. Haber utilizado OWL para modelar el dominio ODP resulta interesante para poder utilizar distintos lenguajes de formalización de comportamiento, en especial si éstos están escritos en OWL también, como es el caso de OWL-S.

La arquitectura diseñada, mostrada en la figura 58, incluye los siguientes componentes:

- Cliente: representa el componente que quiere localizar una entidad, una persona por ejemplo, en la federación.
- PMI: tiene toda la lógica para la reconciliación de instancias y para mantener el identificador de instancia durante una sesión.
- Correlación: sirve como repositorio de perfiles genéricos y podría implementar también la interfaz para obtener el perfil genérico federado; aunque en este caso se deja esta interfaz en el PMI.
- PIDs locales: implementan todas las interfaces autónomas (básicas y complejas) de un sistema.

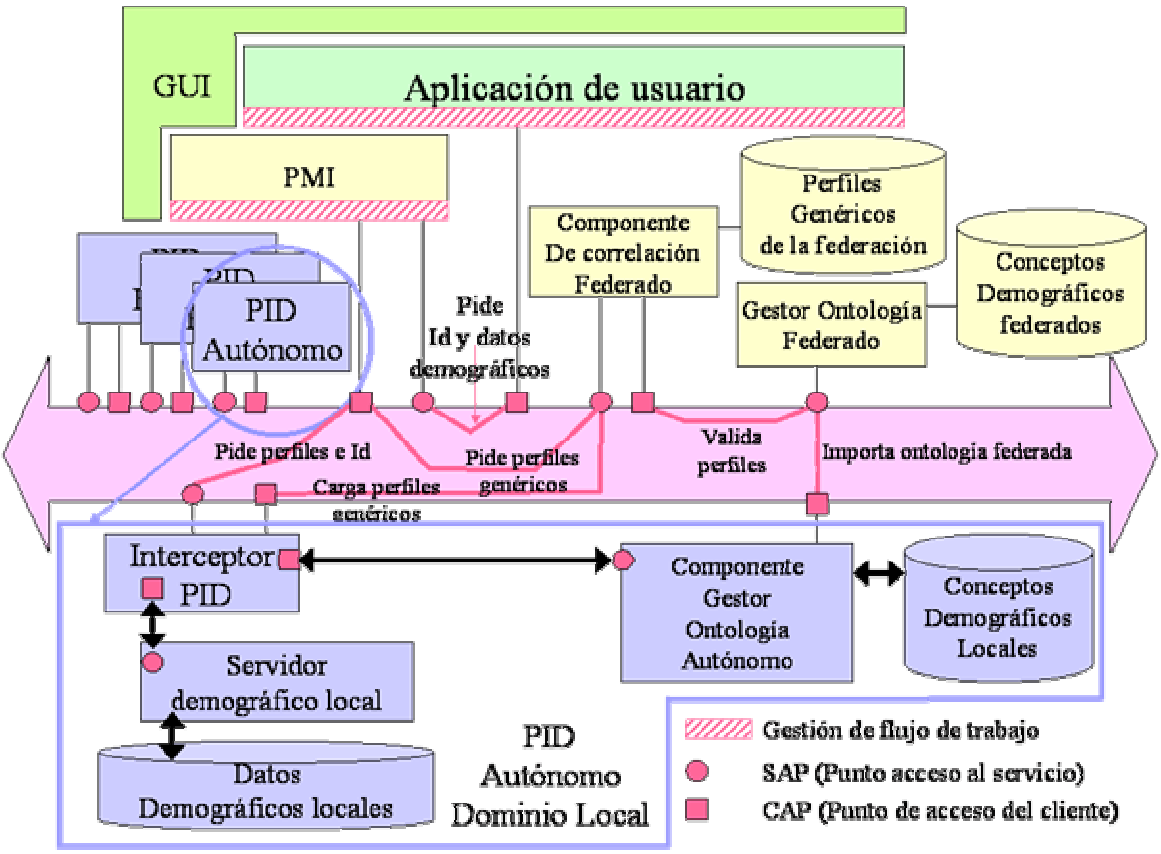


Figura 58. Arquitectura diseñada para la gestión federada de identificadores y datos demográficos

Es importante destacar que la asignación de interfaces a componentes no tiene por qué estar limitada a la aquí presentada, lo realmente importante es la funcionalidad de cada interfaz y los procedimientos para lograr un objetivo determinado. La arquitectura incluye componentes de gestión de ontologías (relacionados con las actividades de gestión semántica) que se encargan de resolver los conflictos semánticos de los datos demográficos; estos componentes no pertenecerían propiamente a las actividades de gestión semántica y podrían ser utilizados en otros grupos de actividades.

Se va a dividir el comportamiento de los componentes implicados en la identificación y recuperación de entidades y perfiles en una federación en tres partes:

- Gestión de perfiles genéricos
- Identificación unívoca de la persona en la federación
- Asentamiento del candidato y obtención del perfil.

Gestión de perfiles genéricos

Si la búsqueda de entidades en la federación se realiza con los atributos de los perfiles genéricos de cada dominio, el procedimiento será óptimo. Para eso es necesario conocer el perfil federado, es decir el perfil que identificaría unívocamente a una entidad en todos y cada uno de los dominios. Para encontrar este perfil se utilizará el método **`Detail Federated Generic Profile`**, que en este caso de uso implementa el componente PMI, como muestra la figura 59. El componente de correlación almacena los perfiles genéricos de todos los sistemas de la federación que lo hayan registrado, de manera que el PMI se encarga de interrogarle pidiéndole los perfiles de los dominios sobre los que se quiere hacer la búsqueda. Antes de devolver los perfiles genéricos solicitados al PMI, el agente de correlación se asegura de que conoce todos los perfiles genéricos de los dominios de interés interrogando a aquellos que no los tuvieran registrados. El PMI analiza los perfiles genéricos y obtiene el perfil genérico federado que devuelve al cliente. En este caso de uso es el PMI el que tiene la lógica para calcular el perfil genérico federado, pero esta funcionalidad podría haberse incluido también dentro del agente de correlación.

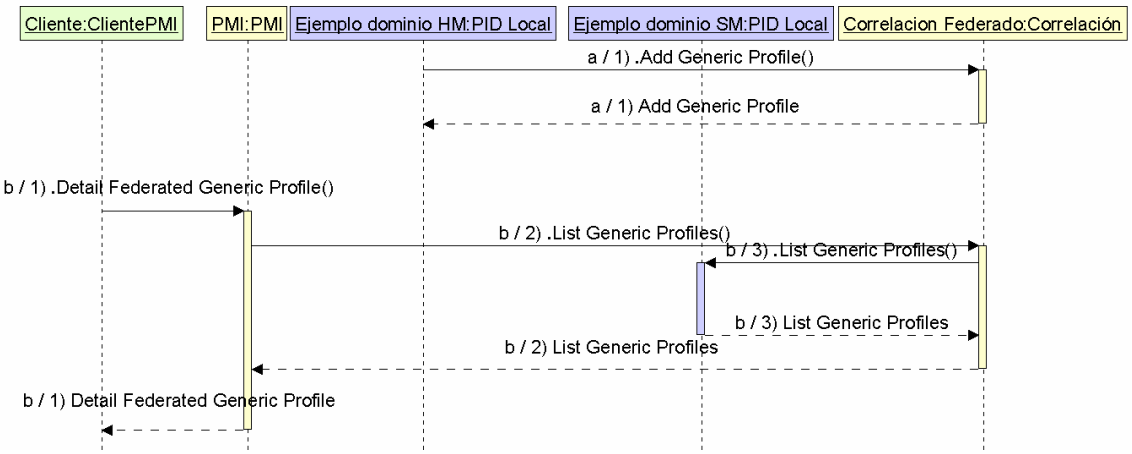


Figura 59. Procedimiento para la obtención del perfil genérico federado

Identificación de la persona (entidad) en la federación

El siguiente paso es comenzar la localización de la entidad, en este caso de la persona. El método **`List Candidates`** iniciará este proceso durante el cual el PMI ejecuta toda la lógica de identificación. En la invocación de **`List Candidates`** se pasan todos los atributos del perfil genérico conocidos por el usuario, aparte de otros atributos conocidos que se quieran usar para la búsqueda.

El procedimiento de búsqueda se muestra en la figura 60 y se puede resumir del siguiente modo:

- **Bucle de búsqueda de candidatos con Hard Matching**
  - **HM directo:** el agente PMI busca todos los dominios donde pueda hacer HM directo con los atributos incluidos en el perfil de búsqueda. Marca los atributos introducidos por el usuario con confianza 1: son definitivamente datos de la persona buscada.
  - **Realizar búsqueda de entidad:** en los dominios en los que esto sea posible se busca la entidad (persona) y se obtendrá uno o cero candidatos.

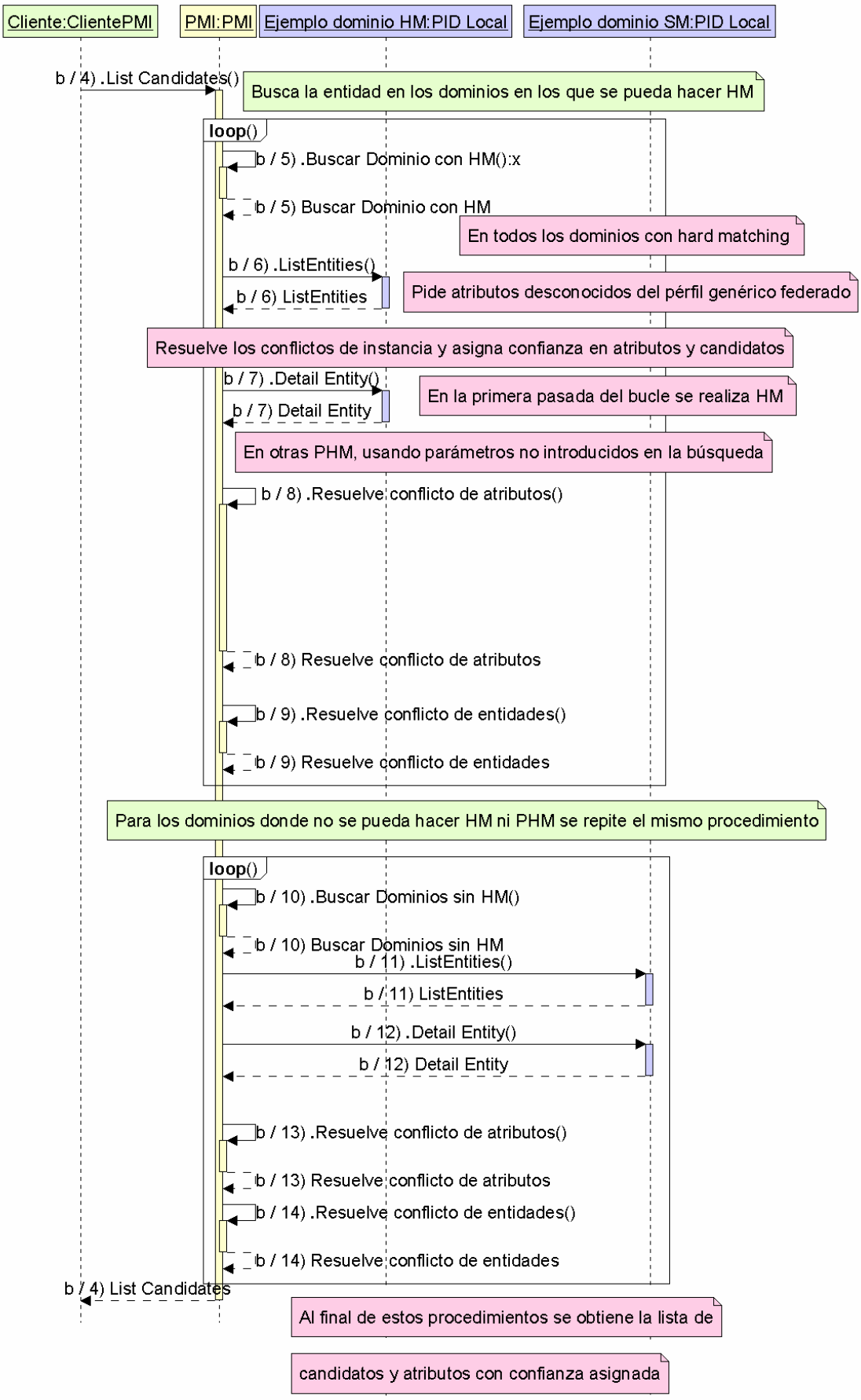


Figura 60. Procedimiento para la identificación de la persona en la federación

- **Obtener atributos provisionales:** en los dominios donde se haya podido hacer un HM directo y se haya devuelto un candidato, se solicitan los atributos necesarios para hacer HM en el resto de dominios.
- **Resolución de conflictos de instancia:** se resuelven los conflictos de atributos y entidades y se marcan los atributos obtenidos con confianza menor que uno (función del resultado de la resolución de conflictos). Se resuelven los conflictos de entidades y se marca la confianza de los candidatos.
- **HM provisional:** con los nuevos atributos se mira dónde podemos hacer **`hard matching`** y se vuelve a repetir el proceso hasta agotar las posibilidades.

– **Bucle de búsqueda de candidatos con Soft Matching**

- **SM:** se buscan los dominios en los que aún no se ha hecho HM ni PHM
- **Realizar búsqueda de entidad:** en cada uno de los dominios encontrados.
- **Resolución de conflictos de instancia:** se resuelven los conflictos de atributos y entidades y se marcan los atributos obtenidos con confianza menor que uno (función del resultado de la resolución de conflictos). Se resuelven los conflictos de entidades y se marca la confianza de los candidatos.

Asentamiento de candidato y obtención de perfil

Una vez seleccionado el candidato, normalmente el de mayor confianza, se envía al PMI con el perfil validado por el usuario. Con esto el PMI asienta la selección, considerando los atributos verificados con confianza 1 y asignando un identificador de entidad que se mantendrá durante toda la sesión y se gestionará la correspondencia con los identificadores de la persona en cada dominio.

El procedimiento se muestra en la figura 61 y se puede resumir como:

- **Verificar atributos:** se recibe el candidato seleccionado y los atributos validados para asentar la persona buscada en esa sesión.

- **Mantener Entity Id en la sesión:** se gestionan las equivalencias del identificador asignado temporalmente al identificador en cada dominio.
- **Obtención del perfil deseado:** a partir de ese momento se puede realizar la búsqueda del perfil deseado, ya sea de la persona o de un rol de entidad persona en particular, como el rol de paciente.

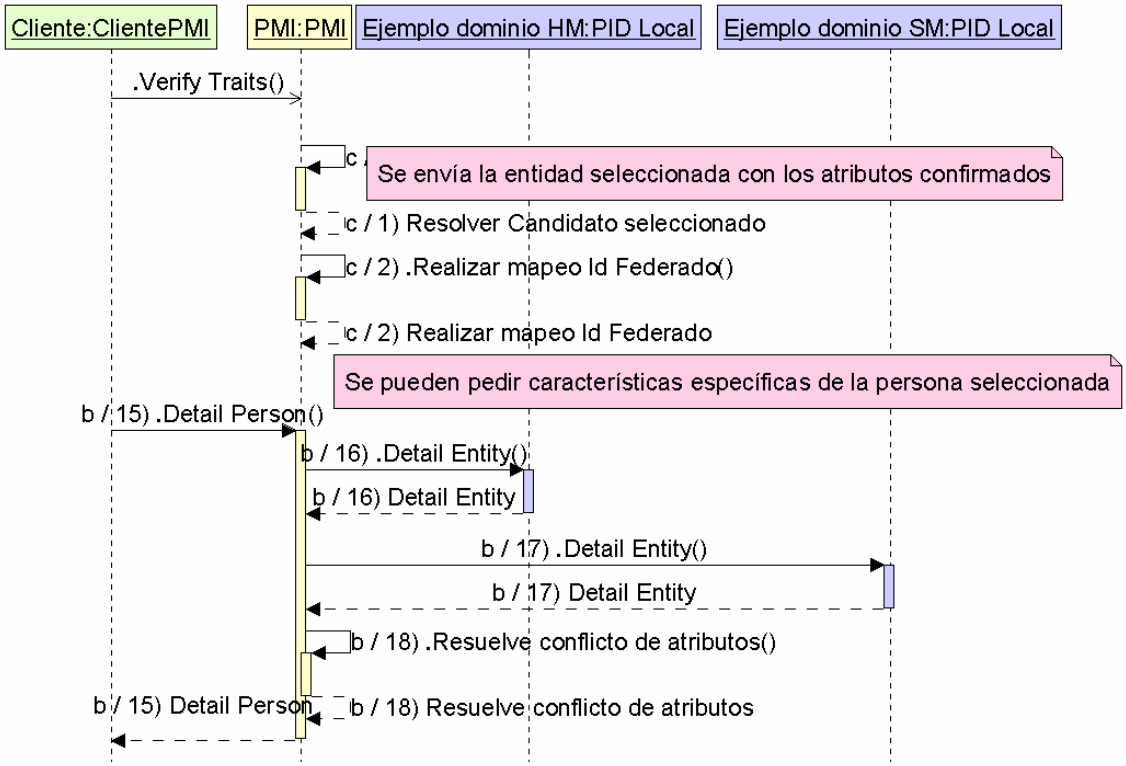


Figura 61. Asentamiento de candidato y obtención de perfil o rol de entidad

7.4.5. Algunas consideraciones finales

En esta especificación computacional se han omitido relaciones con otros grupos de actividades que sería imprescindible considerar a la hora de integrar las tareas aquí presentadas en la organización sanitaria. El modelo de federación elegido en la organización condicionará el modo en el que se lleven a cabo estas interacciones.

Relaciones con las actividades de seguridad

Para poder establecer vínculos en cualquiera de las interfaces presentadas y poder usar las prestaciones ofrecidas en las mismas es imprescindible que tanto el componente que pretende vincularse como el actor al que éste representa tengan los permisos necesarios para ello. Estos requisitos deben ser contemplados en el contrato de entorno.

Resulta interesante destacar que en el modelo de comportamiento mostrado sólo se han considerado tareas de consulta. Sin embargo algunas de las conclusiones obtenidas durante los procedimientos de reconciliación de instancias, y dado el esfuerzo computacional que éstos conllevan, podrían resultar en modificaciones, por parte del PMI, de la entidad de interés o de sus perfiles y atributos en alguno de los dominios. Esto permitiría que la información almacenada en los dominios de la federación estén más actualizados y existan menos inconsistencias, aunque por supuesto las tareas de seguridad se vuelven más complejas.

Es importante no comprometer la seguridad de los datos personales y asegurar el registro de actividades realizadas sobre éstos con propósitos de trazabilidad y auditoría. De manera que los accesos de cualquier tipo a entidades de la clase **`persona`**, y a sus perfiles y atributos asociados, deberá ser registrado en cumplimiento de las leyes aplicables a cada dominio.

Relaciones con las actividades de semántica

Se han presentado la ontología base y la del dominio federadas. Ésta última estará gestionada a través de los componentes de las actividades de semántica que permitirán extenderla o modificarla. De manera que se recurrirá a los componentes de semántica cuando sea necesario realizar tareas, como la consulta, introducción o conversión de conceptos. Así en la figura 58 puede apreciarse que especialmente los interceptores necesitarán interactuar con los componentes semánticos ya que deberán traducir los conceptos locales a los federados y viceversa. Dado que estos componentes no son propiamente del modelo demográfico no se detallan sus interfaces.

Capítulo 8. Contribuciones originales

En este capítulo se enumeran las contribuciones originales aportadas en esta Tesis Doctoral. Es necesario para ello comenzar señalando que la premisa de la misma ha sido considerar que una arquitectura para la integración de sistemas de información y comunicaciones en el entorno sanitario debe ser entendida como una solución abierta y dinámica, que busque siempre facilitar la incorporación de sistemas, siguiendo el paradigma de federación, y el desarrollo de nuevos flujos de cooperación entre componentes que permitan el despliegue de aplicaciones más avanzadas reutilizando las prestaciones ya existentes en la arquitectura.

Estas aplicaciones permitirán combinar y explotar eficientemente el conocimiento almacenado en los sistemas de la organización ofreciendo nuevas facilidades en cualquiera de los aspectos de interés de la organización sanitaria: asistencial, administrativo, de investigación, formativo, etc.

La investigación de esta Tesis Doctoral se ha centrado pues en los requisitos y dificultades de la integración de sistemas de información sanitarios, y en particular en la problemática inherente a la integración según el paradigma de federación. El objetivo era diseñar una metodología de integración que permitiera una solución completamente abierta y escalable desde los dos puntos de vista antes señalados:

- La incorporación de nuevos dominios a la organización

- La incorporación de aplicaciones más avanzadas que reutilizaran eficazmente los componentes disponibles en la arquitectura.

Dado que el gran esfuerzo dedicado por numerosos grupos de investigación y organismos de normalización ha dado como fruto la especificación de estándares para la gestión de la historia clínica electrónica, muy próximos a la madurez, éstos pueden resultar una base importante en el desarrollo de las futuras soluciones de integración de TIC en el seno de la organización sanitaria y han sido especialmente considerados durante el desarrollo de esta Tesis Doctoral. Sin embargo en ningún momento el

objetivo fundamental ha sido la investigación sobre sistemas de historia clínica electrónica en sí mismos, sino la integración de éstos a la arquitectura de la organización sanitaria persiguiendo la explotación eficiente del conocimiento almacenado en el conjunto de sistemas integrados.

La Tesis está enmarcada en el contexto europeo por lo cual se ha estudiado el modelo de arquitectura propuesto en la prenorma del CEN prEN 12967. El análisis detallado de esta prenorma ha dado lugar a la propuesta de una serie de extensiones que contribuyen a facilitar las tareas de integración en entornos federados para conseguir una arquitectura más flexible y eficaz. Todas estas aportaciones se realizan siguiendo el modelo ODP, tal y como la citada prenorma indica, y por tanto se presentan como extensiones a las vistas de empresa, de información y computacional. Estas extensiones se pueden resumir como:

1) Desarrollo del paradigma estratégico para una arquitectura genérica de sistemas de información sanitarios. Las aportaciones más importantes en este sentido han sido:

- Incluir la definición formal de términos relacionados con la metodología ODP y que son imprescindibles a la hora de realizar un correcto diseño. Por ello es conveniente incluirlos en el paradigma estratégico.

- Detallar los principales requisitos de un modelo de integración basado en la federación

- Identificar el papel jugado por los sistemas de historia clínica electrónica dentro de la arquitectura. Dado que estos sistemas acaparan gran parte de la atención de científicos y desarrolladores es imprescindible identificar perfectamente su posición en el seno de la arquitectura.

2) Modificaciones y extensiones a la vista de empresa.

- Reestructuración de la clasificación de actividades a realizar en una organización sanitaria, ampliando el número de tareas básicas implicadas, especialmente en cuanto a gestión de seguridad, comunicaciones y semántica.

- Presentación detallada de los objetivos de cada flujo de trabajo y grupo de actividades considerando la nueva clasificación realizada.

- Mención especial requiere la extensión de los requisitos impuestos a la arquitectura para la gestión de semántica. En este punto se incluyen actividades de gestión de ontologías, reglas e inferencias, que no están consideradas en la prenorma y que facilitan la integración basada en la federación.

3) Modificación del paradigma de modelado de la vista de información y extensiones a los modelos de información.

- Se considera el modelo dual en el diseño de la vista de información que, de este modo, se desarrolla según un modelo ontológico de dos niveles: una ontología base estática y una ontología dinámica, más cercana al dominio, y fácilmente gestionable a través de componentes específicos de la arquitectura, que son presentados en las distintas vistas de la arquitectura.

- Inclusión en la ontología base de los conceptos necesarios para la aplicación de este paradigma de modelado. Éstos, junto al uso de técnicas de gestión de ontologías, facilitan la extensión o modificación del modelo de información dinámico a medida que los conceptos o la cobertura del dominio cambien, así como la integración de nuevos sistemas a la arquitectura.

- Desarrollo de las ontologías base para cada flujo de trabajo y actividad, considerando las modificaciones de clasificación propuestas. Al mismo tiempo se presentan distintas normas que cubren diversos aspectos tratados en HISA y que deberán ser tenidas en cuenta.

4) Modificaciones y extensiones de la vista computacional:

- La mayor aportación a la vista computacional es la propuesta de un lenguaje abierto para la especificación de componentes que cubra todas las necesidades de especificación de una arquitectura basada en ODP como es la prenorma 12967. Esta ontología está escrita en OWL lo que permite la especificación de componentes y servicios utilizando algún editor de ontologías como Protègè. Al

dar dentro de esta vista un fuerte peso al desarrollo de un lenguaje computacional suficientemente abierto se facilita la integración de sistemas siguiendo el paradigma de federación. Por otro lado se favorece la adaptación de la vista computacional a las vistas de ingeniería y tecnológica propias de cada implementación.

- Se presenta una extensión a los componentes básicos que se deberían encontrar en cualquier arquitectura sanitaria. Especialmente se extienden las funciones relacionadas con la gestión de ontologías que soportarán, entre otras, las actividades de desarrollo de la ontología dinámica del dominio o las de traducción desde ontologías locales a la federada.

5) Validación de la metodología propuesta: desarrollo de las actividades relativas a la gestión demográfica.

- Desarrollo del paradigma estratégico: Se detallan los requisitos y dificultades de las actividades de gestión demográfica. Se presta una atención especial a las tareas de identificación de personas y de recuperación de datos personales en un entorno completamente federado. Estas actividades no se consideran propias del flujo de trabajo del sujeto de cuidado ya que se habla de la identificación de personas en general y no de pacientes. De manera que en realidad serán reutilizadas en varios flujos de trabajo y actividades de la arquitectura sanitaria.

- Vista de empresa: Se presentan las actividades propias de los componentes demográficos de la arquitectura sanitaria y su relación con los flujos de trabajo y actividades de la organización sanitaria, identificando especialmente necesidades comunes y relativas a la identificación de personas.

- Vista de información: Se desarrolla en detalle la ontología base. Esta ontología incluye especialmente elementos del modelo de información para el flujo del sujeto de cuidado, del modelo de seguridad y del de gestión recursos, extendiéndolos con elementos que facilitan la reutilización en todos ellos y el desarrollo de las interfaces para la gestión demográfica.

- Vista computacional: Se diseña la arquitectura de componentes demográficos encargados de la identificación y recuperación de datos de personas en la federación siguiendo un modelo completamente federado que no necesita la centralización de información demográfica ni de identificadores. Se utiliza la ontología ODP desarrollada para especificar los componentes de la arquitectura. Se detalla el flujo de trabajo para la identificación unívoca de una persona en la federación y la obtención de sus datos personales.

- El diseño realizado podrá incorporarse fácilmente a la arquitectura de sistemas sanitarios para ser reutilizado dentro de la misma y servir de soporte a aplicaciones de más alto nivel.

Capítulo 9. Publicaciones originales

La investigación realizada durante el desarrollo de esta Tesis doctoral ha dado lugar a las siguientes publicaciones

9.1. Año 2006

9.1.1. Publicaciones en revistas

Demographic management in a federated healthcare environment

Román, L.M. Roa, J. Reina and G. Madinabeitia

International Journal of Medical Informatics

Aceptado y pendiente de publicación.

9.2. Año 2005

9.2.1. Capítulos de libro

Standardization of Demographic Service for a Federated Healthcare Environment

Román, L.M. Roa, L.J. Reina and G. Madinabeitia

Medical and Care Compunetics 2

Editor: L.Bos et al.

ISBN: 1-58603-520-7

ISSN: 0926-9630

pág 259-269

9.2.2. Actas de congresos

Internacionales

Open Laboratory Information System Based On CEN Data Types And LOINC Codes

I.Román, L.M. Roa, G. Madinabeitia and R. García

Ontology Server for the integration of Demographic Servers

I.Román, L.M. Roa, G. Madinabeitia and A.M. Oria

3rd European Medical & Biological Engineering Conference IFMBE European Conference on Biomedical Engineering

Generación y acceso al conocimiento en telemedicina: nuevos enfoques y soluciones.

M. Prado; L. M.Roa; L. J. Reina; I. Román

Conferencia Iberoamericana en Sistemas, Cibernética e Informática (CISCI 2005)

Nacionales

Servicio de registro de actividad en sistemas de información sanitarios.

I. Román, G. Madinabeitia, L.M. Roa, D.D. García

ISBN: 84-7402-325-4

pág 259-262

XXIII Congreso anual de la Sociedad Española de Ingeniería Biomédica. (CASEIB 2005., Madrid)

9.3. Año 2004

9.3.1. Actas de congresos

Internacionales

A Semantic Web architecture for the integration of Electronic Healthcare Record

Román, I., Roa, L., Vozmediano, J.M.

MEDICON and HEALTH TELEMATICS 2004, X Mediterranean Conference on Medical and Biological Engineering

IFMBE Proceedings. Volume 6, 2004

ISSN: 1727-1983

ISBN: 88-7780-308-8

Nacionales

Desarrollo de un servidor COAS con tecnología de servicios web.

J. Álamo, I. Román, L.M. Roa, G. Madinabeitia

Consideraciones sobre el control de acceso a recursos en una arquitectura de historia clínica.

I. Román, C. Paz, G. Madinabeitia, L.M. Roa

XXII Congreso anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2004, Santiago de Compostela)

ISBN: 84-688-9318-8

9.4. Año 2003

9.4.1. Actas de congresos

Internacionales

Management of Discharge Summaries based on XML

Proc of the 4th Annual IEEE Conf on Information Technology Applications in Biomedicine, UK (ITAB, 2003)

I. Román, L.M. Roa, J.M. Vozmediano,L. Corrales

IEEE Catalog Number: 03TH8655C

ISBN: 0-7803-7668-4

Nacionales

Consideraciones metodológicas para conseguir la historia clínica virtual como una composición de servicios

I. Román, L.M. Roa, J.M. Vozmediano

Desarrollo de un servidor LQS y un cliente con interfaz web de acceso

I. Román, L.M. Roa, J.M. Hormigo, J.M. Vozmediano

Seguridad en el intercambio de documentos médicos en formato xml mediante los mecanismos de firma digital y encriptado siguiendo la normativa europea.

A. Martinez, I. Román, L.M. Roa

XXI Congreso anual de la sociedad Española de ingeniería biomédica. (CASEIB 2003, Mérida)

ISBN:84-95454-62-9

9.5. Año 2002

9.5.1. Actas de congresos

Nacionales

Planificación de base de datos e interfaz web como primer paso a la historia clínica electrónica

Román, I., Roa, L., Vozmediano, J.M., Ramos, E.

XX Congreso Anual. Sociedad Española de Ingeniería Biomédica. (CASEIB 2002, Zaragoza)

ISBN/D.L.: 84-600-9818-4

9.6. Año 2001

9.6.1. Actas de congresos

Nacionales

NefroAlt: Una aplicación para la realización del informe de alta

Román, I., Roa, L., Palma, A., Milán, J.A.

XIX Congreso Anual. Sociedad Española de Ingeniería Biomédica. (CASEIB 2001, Madrid).

Capítulo 10. Línea de investigación

La investigación realizada durante el desarrollo de esta Tesis Doctoral se ha enmarcado principalmente en el contexto del proyecto MEDICI (Monitorización Extrahospitalaria para el Diagnóstico y la Investigación de la Cardiopatía Isquémica. Herramientas para la Adquisición, Visualización, Integración y Descubrimiento de Conocimiento Médico, MCYT TIC2003-09300-C04). La misión del grupo de trabajo del doctorando dentro de este proyecto era el diseño de la metodología de integración de los distintos sistemas que conformaban la arquitectura desarrollada, con la finalidad de facilitar la explotación del conocimiento clínico implícito en los mismos y la generación de nuevo conocimiento.

De manera que esta Tesis Doctoral supone el punto de partida de la carrera investigadora de la doctoranda que, en líneas generales, estará orientada principalmente a aportar soluciones de integración de sistemas de información sanitarios. Los métodos y modelos desarrollados deberán siempre perseguir la reutilización eficiente de componentes y prestaciones para permitir una explotación adecuada del conocimiento almacenado en los sistemas de la organización, así como el desarrollo de nuevas aplicaciones que faciliten la generación de conocimiento.

En esta Tesis Doctoral se presenta la base metodológica para el diseño de una arquitectura de integración. Las futuras aportaciones en esta línea deberán ser conformes a la misma, tal y como se ha hecho para las tareas demográficas.

Ante el convencimiento de que el avance en las técnicas de gestión semántica facilitará el diseño de mecanismos de integración flexibles que permitan la federación de sistemas desarrollados independientemente y para fines diversos, de forma sencilla y eficiente, la investigación considerará continuamente los avances realizados en este campo y, en la medida de lo posible, realizará aportaciones a los mismos.

Como trabajo más inmediato la investigación comenzada en el proyecto MEDICI será continuada, D.m., en el proyecto IDEATIO (Herramientas y Dispositivos Inteligentes para el Control de la Calidad Asistencial en Procesos de Seguimiento Integrado

Intra/Extrahospitalario de pacientes, TIN2006-15460-C04-03), que actualmente se encuentra pendiente de revisión.

El objetivo principal del grupo de trabajo al que pertenece el doctorando dentro de este proyecto será el diseño de métodos y modelos para la integración de sistemas de gestión de Guías de Práctica Clínica (GPC) a la arquitectura de sistemas de información y comunicaciones de una organización sanitaria. La finalidad es por un lado facilitar al clínico la elaboración del plan de cuidado personalizado más adecuado al paciente y por otro enriquecer la validación, revisión y creación de GPC. Estas mejoras se pretenden obtener gracias a la óptima explotación del conocimiento clínico recogido en el conjunto de sistemas que forman parte de la organización sanitaria, en especial los de HCE, y al correcto establecimiento de los vínculos entre éste y la gestión de las GPC.

De este modo la línea de investigación más inmediata podría describirse como el desarrollo de modelos de integración de los sistemas de HCE y los de gestión de GPC, que son entendidos como el mecanismo fundamental de la práctica de la medicina basada en la evidencia. Dichos modelos estarán desarrollados en el contexto metodológico propuesto en esta Tesis Doctoral.

A medio y largo plazo es de especial interés el desarrollo de algunos de los flujos de trabajo y actividades presentados, en particular:

- Desarrollar las actividades que en esta Tesis se han denominado “de gestión de comunicaciones”. Profundizando en los mecanismos de descubrimiento y selección de interfaces y componentes, así como de establecimiento de vinculaciones, considerando tanto el objetivo final como los requisitos de calidad de servicio. Esta línea de investigación deberá considerar distintos escenarios del contexto sanitario, telemedicina, telemonitorización, entorno hospitalario, etc, y proporcionar métodos abiertos que cubran todos ellos.
- Profundizar en las tareas de reconciliación de atributos e instancias imprescindibles en entornos federados. Analizar los algoritmos existentes de comparación de valores de atributos y proponer mejoras a los mismos y particularizaciones para los tipos de datos utilizados en el dominio sanitario.
- El desarrollo de métodos para integrar en la organización sanitaria aplicaciones avanzadas reutilizando componentes ya existentes en la misma y definiendo

formalmente mecanismos de descripción de flujo de trabajo y de interacción entre agentes de la arquitectura.

En general, la línea de investigación propuesta podría resumirse como el desarrollo de métodos y modelos para la integración de componentes, procesos y aplicaciones en el entorno sanitario. Considerando especialmente que una estrategia correcta de desarrollo, implantación e integración de sistemas es el soporte fundamental para el diseño de aplicaciones avanzadas que permitan la extracción, generación y explotación eficientes del conocimiento almacenado en la organización. El foco de interés especial será el conocimiento clínico que permita la mejora de la calidad asistencial al paciente, pero sin perder de vista en ningún momento las conexiones y sinergias con el resto de actividades de la organización.

Bibliografía

[1] Peter G.Goldshmidt, "HIT and MIS: Implications of health information technology and medical information systems" Communications of the ACM vol. 48, no. 10, pp. 69-74, Oct.2005.

[2] Lisa Pizziferri, Anne F.Kittler, Lynn A.Volk, Melissa M.Honour, Sameer Gupta, Samuel Wang, Tiffany Wang, Margaret Lippincott, Qi Li, and David W.Bates, "Primary care physician time utilization before and after implementation of an electronic health record: A time-motion study" Journal of Biomedical Informatics vol. 38, pp. 176-188, 2005.

[3] Sotiris A.Pavlopoulos and Anastasios N.Delopoulos, "Designing and implementing the transition to a fully digital hospital" IEEE Transactions on information technology in biomedicine vol. 3, no. 1 Mar.1999.

[4] Horsch A. and Balbach T., "Telemedical Information Systems" IEEE Transactions on information technology in biomedicine vol. 3, no. 3 Sept.1999.

[5] M. Sobol, Alverson M., and Lei D., "Barriers to the adoption of computerized technology in health care systems" Topic in Health Information Management vol. 19, no. 4 May1999.

[6] Thomas Beale, A.Goodchild, and Sam Heard, "Design Principles for the EHR. Revision 2.4,"May2002.

[7] CEN/ISSS e-Health Standardization Focus Group, "Current and future standardization issues in the e-Health domain: Achieving interoperability (Final Version),"Mar.2005.

[8] CEN /TC 251.Secretariat: SIS, "Health Informatics - Service Architecture (HISA). prEN 12967-1 Part 1: Enterprise viewpoint" 2005.

[9] Indra y Junta de Andalucía, "DIRAYA" 2005.
URL:http://www.juntadeandalucia.es/repositorio/projects/sas-diraya/
Nota: Último acceso enero 2006

[10] Susanne Busse, Ralf-Detlef Kutsche, Ulf Leser, and Herbert Weber, "Federated Information Systems: Concepts, Terminology and Architectures,"Forschungsberichte des Fachbereichs Informatik Bericht Nr. 99-9, Apr.1999.

[11] "Proyecto CLEF" 2006.
URL:www.clinical-escience.org/more.html
Nota: Último acceso julio de 2005

[12] Alan Rector, Adel Taweel, Jeremy Rogers, David Ingram, Dipak Kalra, Robert Gaizauskas, Mark Hepple, Jo Milan, Richard Powers, Donia Scott, and Peter Singleton, "Joining up Health and BioInformatics: e-Science meets e-Health" 2005.

[13] Alan Rector, Jeremy Rogers, Adel Taweel, David Ingram, Dipak Kalra, Jo Milan, Peter Singleton, Robert Gaizauskas, Mark Hepple, Donia Scott, and Richard Powers, "CLEF - Joining up Healthcare with Clinical and Post-Genomic Research" 2005.

[14] J. H. van Bemmel, E. M. van Mulligen, B. Mons, M. van Wijk, J. A. Kors, and J. van der Lei, "Databases for knowledge discovery: Examples from biomedicine and health care" International Journal of Medical Informatics vol. In Press, Corrected Proof.

[15] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento
distribuido abierto: X.901 Modelo de referencia: Visión de conjunto" 1997.

[16] Ee-Peng Lim, Jaideep Srivastava, and Shashi Shekhar, "Resolving attribute incompatibility in database integration: an evidential reasoning approach" IEEE Computer Society Ninth International Conference on Data Engineering 1994,IEEE Computer Society Ninth International Conference on Data Engineering, pp. 154-163.

[17] Debabrata Dey and Sumit Sarkar, "A Distance-Based Approach to Entity Reconciliation in Heterogeneous Databases", 14 ed 2002, pp. 567-582.

[18] Ee-Peng Lim and Jaideep Srivastava, "Entity Identification in Database Integration: An Evidential Reasoning Approach" Int'l Conf. Next Generation Database 1993,Proc.Int'l Conf.Next Generation Database.

[19] Ee-Peng Lim and Jaideep Srivastava, "Entity Identification in Database Integration" *IEEE Computer Society Ninth International Conference on Data Engineering* 1993,IEEE Computer Society Ninth International Conference on Data Engineering, pp. 19-23.

[20] Huimin Zhao and Sudha Ram, "Entity identification for heterogeneous database integration—a multiple classifier system approach and empirical evaluation" Information Systems vol. 30, pp. 119-132, 2005.

[21] Ee-Peng Lim and Roger H.L.Chiang, "Accommodating instance heterogeneities in database integration" Decision Support Systems vol. 38 2004.

[22] Jeffrey Parsons and Yair Wand, "Attribute-Based Semantic Reconciliation of Multiple Data Sources" Journal of Data Semantics, pp. 21-47, 2003.

[23] Javier Carnicero Giménez de Azcárate and José Manuel Vázquez López, "La identificación, un requisito previo a la historia de salud electrónica,"V (Capítulo 4), Dec.2003.
Nota: Último acceso diciembre 2005

[24] Werner Ceusters and Barry Smith, "Strategies for referent tracking in electronic health records" Journal of Biomedical Informatics vol. Pendiente de publicación Sept.2005.

[25] Thomas Beale, "The GEHR Archetype System. (The Good Electronic Health Record Project)" 2005.
URL:www.gehr.org/technical/archetypes/gehr_archetypes.html
Nota: Último acceso enero 2005

[26] openEHR Foundation, "The archetype object model. Revision: 0.5.1" 2005.

[27] CEN /TC 251.Secretariat: SIS, "Health informatics - Electronic health record communication. prEN 13606-2 Part 2: Archetypes (Version 2.2)" 2005.

[28] CEN /TC 251.Secretariat: SIS, "Health informatics - Electronic health record communication.PrEN13606-3 Part 3: Reference archetypes and term lists" 2005.

[29] OMG CORBAmed DTF, "Person Identification Service (PIDS) Specification. Version 1.1." 2001.

[30] G.Thimm, S.Ben-Yacoub, and J.Luettin, "Evaluating the Complexity of Databases for Person Identification and Verification" *CAIP'99* 1999, pp. 49-56.

[31] David W.Forslund, Rebecca K.Smith, and Thomas C.Culpepper, "Federation of the Person Identification Service between Enterprises" *AMIA Symp. 2000* 2000,AMIA Symp, pp. 240-244.

[32] Sara Facchinetti, Damon Berry, Sebastien Pardon, and Jane Grimson, "Distributed Patient Records require Distributed Patient ID Management: Enter PIDS? " *PICNIC* 2002,Medilink Programme Seminar Series.
Nota: Último acceso diciembre 2005

[33] openEHR Foundation, "The openEHR Demographic Information Model. Revision 2.0" 2006.

[34] CEN /TC 251.Secretariat: SIS, "Health informatics- General purpose components for messages- Part2: Non clinical prEN 14822-2:2003 (E)" 2003.

[35] José Luis Monteagudo Peña and Carlos Hernández Salvador, "Estándares para la historia clínica electrónica,"V (Capítulo 7), Dec.2003.
Nota: Último acceso diciembre 2005

[36] "Página web del comité técnico 251 del CEN".
URL:www.centc251.org
Nota: Último acceso diciembre 2005

[37] "Página web del CEN".
URL:www.cenorm.be
Nota: Último acceso enero 2005

[38] "Página web del CENELEC".
URL:www.cenelec.org
Nota: Último acceso 2003

[39] CEN /TC 251.Secretariat: SIS, "Health informatics - Electronic health record communication. prEN 13606-1 Part 1: Reference model" 2004.

[40] CEN /TC 251.Secretariat: SIS, "Health informatics - Electronic health record communication.prEN 13606-4 Part 4: Security requirements and distribution rules" 2005.

[41] CEN /TC 251.Secretariat: SIS, "Health informatics - Electronic health record communication.prEN 13606-5 Part 5: Messages for exchange" 2004.

[42] CEN /TC 251.Secretariat: SIS, "Health informatics- General purpose information components- Part1: Overview. prEN 14822-1:2003 (E)" 2003.

[43] CEN /TC 251.Secretariat: SIS, "Health informatics- General purpose components for messages- Part3: Clinical. prEN 14822-3:2003 (E)" 2003.

[44] CEN /TC 251.Secretariat: SIS, "Health Informatics - Service Architecture (HISA).prEN 12967-2 Part 2: Information viewpoint" 2005.

[45] CEN /TC 251.Secretariat: SIS, "Health Informatics - Service Architecture (HISA). prEN 12967-3 Part 3: Computational viewpoint" 2005.

[46] "Página web de ISO".
URL:www.iso.ch/iso/en/stdsdevelpment
Nota: Último acceso 2003

[47] "HL7 (Página Web)".
URL:www.hl7.org
Nota: Último acceso diciembre 2003

[48] "Página web de la ACR-NEMA".
URL:medical.nema.org
Nota: Último acceso 2003

[49] National Electrical Manufactures Association, "Digital Imaging and Communications in Medicine (DICOM) Part 1: Introduction and Overview Final Draft: PS 3.1-200".
URL:global.ihs.com

[50] "DICOM. STRATEGIC DOCUMENT. Version 5.1,"Dec.2005.
Notes: Último acceso enero 2006

[51] "Digital Imaging and Communications in Medicine (DICOM). Part 1: Introduction and Overview,"2004.
Nota: Último acceso enero 2006

[52] OMG, "Normas del Healthcare Domain Task Force (página web)".
URL:http://healthcare.omg.org/Roadmap/corbamed_roadmap.htm
Nota: Último acceso diciembre 2005

[53] OMG CORBAmed DTF, "Lexicon Query Service (LQS)" 1998.

[54] OMG CORBAmed DTF, "Clinical Observations Access Service (COAS) Revised submission" 1998.

[55] OMG CORBAmed DTF, "Clinical Image Access Service (CIAS) Revised Submission" 2000.

[56] OMG CORBAmed DTF, "Resource Access Decision Facility Specification (RAD). Version 1.0" 2001.

[57] "PICNIC System Architecture ,"**D 2.4**, Mar.2003.

[58] J. Grimson, W. Grimson, D. Berry, and G. Stephens, "A CORBA Based Integration of Distributed Electronic Healthcare Records Using the Synapses Approach" IEEE Transactions on information technology in biomedicine vol. 2, no. 3 Aug.1998.

[59] Kalra D., "The Synapses Project: Final Deliverable. 10 chapters, 64 pages." Synapses ODP Information Viewpoint. EU Telematics Application Programme, Brussels, 1998.

[60] M. Tsiknakis, D. G. Katehakis, and S. C. Orphanoudakis, "An open, component-based information infrastructure for integrated health information networks" International Journal of Medical Informatics vol. 68, no. 1-3, pp. 3-26, Dec.2002.

[61] "Página web de umls".
URL:www.nlm.nih.gov/research/umls/
Nota: Último acceso 2004

[62] "Página web de códigos icd10".
URL:www.who.int/msa/mnh/ems/icd10/icd10.htm
Nota: Último acceso 2004

[63] "Página web SNOMED".
URL:www.snomed.org
Nota: Último acceso 2003

[64] "Página web LOINC".
URL:www.regenstrief.org/loinc/
Nota: Último acceso junio 2005

[65] "Página web EUGLOSS".
URL:allserv.rug.ac.be/rudstich/eugloss/welcome.html
Nota: Último acceso 2003

[66] openEHR Foundation, "The openEHR Common Information Model. Revision: 2.0" 2006.

[67] openEHR Foundation, "The OpenEHR Integration Information Model.Revision 0.5" 2005.

[68] openEHR Foundation, "The *open*EHR Data Structures Information Model. Version 1.6rc2,"Dec.2005.

[69] openEHR Foundation, "The *open*EHR Data Types Information Model. Revision: 2.0rc2,"Dec.2005.

[70] openEHR Foundation, "The *open*EHR EHR_EXTRACT Information Model. Revision: 1.3.5,"Dec.2004.

[71] openEHR Foundation, "The openEHR Archetype System. Revision 0.3.1,"Dec.2003.

[72] Berthold B.Wein, Marco Eichelberg, Alexander Ihls, and Eric Poiseau, "IHE Integrating the Healthcare Enterprise-an update for Information Technology Infrastructure for 2005" International Congress Series 1281, pp. 169-174, 2005.

[73] Teresa Ariza, "Contribucionees a la gestión de dispositivos en sistemas basados en CORBA" 2002.

[74] D.De Roure, N.R.Jennings, and N.R.Shadbolt, "Research Agenda for the Semantic Grid: A Future e-Science Infraestructure" 2001.

[75] "Página web del OMG".
URL:www.omg.org
Nota: Último acceso 2004

[76] Kent Wreder and Yi Deng, "Architecture-Centered Enterprise System Development and Integration Based on Distributed Object Technology Standard".

[77] David W.Forslund, James E.George, Eugene M.Gavrilov, and Torsten A.Staab, "Distributed Telemedicine Using High-Performance Computing Over the Internet" High Performance Distributed Computing June1998.

[78] M. Wooldridge and N. R. Jennings, "Agent Gheories, Architectures and Languages: a Survey" 1995,Workshop on Agent Theories, Architectures and Kanguages, 1995.

[79] V. Della Mea, "Agents Acting and Moving in Healthcare Scenario- A paradigm for Telemedical Collaboration" IEEE Transactions on information technology in biomedicine vol. 5, no. 1 Mar.2001.

[80] M. M. G. DeRico, R. B. J. Byrnes, J. H. Schafer, J. A. Marin, M. D. McNett, and G. F. Stone, "Using intelligent agents to combine heterogeneous distributed data", 3 ed 1998,III IEEE International Conference on Systems, Man, and Cybernetics.

[81] "Página Web de la FIPA".
URL:www.fipa.org
Nota: Último acceso septiembre 2005

[82] Foster I. and Kesselman C., *The Grid: Blueprint for a new computing Infrastructure"* Morgan Kaufmann, 1999.

[83] David De Roure, Nicholas R.Jennings, and Nigel R.Shadbolt, "The Semantic Grid: Past, Present, and Future" Proceedings of the IEEE vol. 93, no. 3, pp. 669-681, Mar.2005.

[84] openEHR Foundation, "The Archetype Definition Language (ADL). Revision: 1.3.1" 2006.

[85] Rector AL, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, and Chris Wroe, "OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns" 2005.

[86] W3C, "OWL Web Ontology Language, Guide" 2004.
URL:www.w3.org/TR/owl-guide/
Nota: Último acceso septiembre 2005

[87] W3C, "Web Ontology Language (OWL) Abstract Syntax and Semantics, W3C Working Draft 3" 2003.
URL:www.w3.org/TR/2003/WD-owl-semantics-20030203/

[88] W3C, "OWL Web Ontology Language. Use Cases and Requirements,"Feb.2004.
Notes: Último acceso diciembre 2005

[89] W3C, "OWL Web Ontology Language. Overview,"Feb.2004.
Notes: Último acceso diciembre 2005

[90] Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities" Scientific American May2001.

[91] Jérôme Euzenat and Amedeo Napoli, "The Semantic Web: Year One" IEEE Intelligent Systems 2003.

[92] David Martin, Mark Burstein, Jerry Hobbs, Ora Lassila, Drew McDermott, Sheila McIIraith, Srini Narayanan, Massimo Paolucci, Bijan Parsia, Terry Payne, Evren Sirin, Naveen Srinivasan, and Katia Sycara, "OWL-S: Semantic Markup for Web Services (Version 1.1)" 2005.
URL:www.daml.org/services/owl-s/1.1/overview
Nota: Último acceso noviembre 2005

[93] Ian Horrocks, Peter F.Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean, "SWRL: A Semantic Web Rule Language. Combining OWL and RuleML" 2004.
URL:www.w3.org/Submission/2004/SUBM-SWRL-20040521
Nota: W3C Member Submission. Último acceso Noviembre 2005

[94] David Martin, Mark Burstein, Ora Lassila, Massimo Paolucci, Terry Payne, and Sheila McIIraith, "Describing Web Services using OWL-S and WSDL,"Oct.2005.

[95] The OWL Services Coalition, "OWL-S: Semantic Markup for Web Services" 2005.
URL:www.daml.org/services/owl-s/1.0/owl-s.html
Nota: Último acceso Octubre 2005

[96] Katia Sycara, Massimo Paolucci, Julien Soudry, and Naveen Srinivasan, "Dinamic Discovery and Coordination of Agent-Based Semantic Web Services" IEEE Internet Computing, pp. 66-73, June2004.

[97] Evren Sirin, James Hendler, and Bijan Parsia, "Semi-automatic Composition of Web Services using Semantic Descriptions" Web Services: Modeling, Architecture and Infrastructure workshop in ICEIS 2003 2003,Web Services: Modeling, Architecture and Infrastructure workshop in ICEIS 2003.

[98] Massimo Paolucci, Takahiro Kawamura, Terry Payne, and Katia Sycara, "Semantic Matching of Web Services Capabilities" 2005.

[99] Massimo Paolucci, Takahiro Kawamura, Terry Payne, and Katia Sycara, "Importing the Semantic Web in UDDI" International Workshop on Web Services, E-Business, and the Semantic Web 2002,Lecture Notes In Computer Science; Vol.2512, pp. 225-236.

[100] OASIS, "UDDI Spec TC V4 Requirement Value sets, taxonomies and semantics" 2005.
URL:www.oasis-open.org/
Nota: Último accesso enero 2005

[101] Chris Peltz, "Web Services Orchestration and Choreography" Computer, pp. 46-52, Oct.2003.

[102] David Geer, "Taking Steps to Secure Web Services" Computer, pp. 14-16, Oct.2003.

[103] Asuman Dogac, Gokce B.Laleci, Serkan Kirbas, Yildiray Kabak, Siyamed S.Sinir, Ali Yildiz, and Yavuz Gurcan, "Artemis: Deploying semantically enriched Web services in the healthcare domain" Information Systems vol. Pendiente de publicación Apr.2005.

[104] Jwan Mella, Andrea Zisman, Dave nurse, and Geroge Spanoudakis, "Using Web Services to Support Legacy Clinical Systems" 2005.

[105] Laura Maruster and René J.Jorna, "From Data to Knowledge: A Method for Modeling Hospital Logistic Processes" IEEE Transactions on information technology in biomedicine vol. 9, no. 2 June2005.

[106] Rector AL, Solomon WD, and Nowlan WA, "A terminology server for Medical Language and Medical Information Systems" IMIA 94 1994,IMIA.

[107] L. F. Soualmia and S. J. Darmoni, "Combining different standards and different approaches for health information retrieval in a quality-controlled gateway" International Journal of Medical Informatics vol. 74, no. 2-4, pp. 141-150, Mar.2005.

[108] P. H. M. Chang, S. Von-Wun, C. Tai-Yu, L. Wei-Shen, S. Shiun-Cheng, and H. Yu-Ling, "Automating the determination of open reading frames in genomic sequences using the Web service techniques - a case study using SARS coronavirus" 2004,Bioinformatics and Bioengineering, 2004.BIBE 2004.Proceedings.Fourth IEEE Symposium on, pp. 451-458.

[109] D.Pérez-Rey, V.Maojo, M.García-Remesal, R.Alonso-Calvo, H.Billhardt, F.Martín-Sánchez, and A.Sousa, "ONTOFUSION:Ontology-based integration of genomic and clinical databases" Computers in Biology and Medicine vol. Pendiente de publicación Sept.2005.

[110] E. K. Neumann, E. Miller, and J. Wilbanks, "What the semantic web could do for the life sciences" Drug Discovery Today: BIOSILICO vol. 2, no. 6, pp. 228-236, Nov.2004.

[111] C. Goble, "Carole Goble discusses the impact of semantic technologies on the life sciences" Drug Discovery Today: BIOSILICO vol. 2, no. 1, pp. 4-6, Jan.2004.

[112] G. Bai and P. Zhang, "Developing a semantic Web services for interoperability in diabetic healthcare", 1 ed 2005,Services Systems and Services Management, 2005.Proceedings of ICSSSM '05.2005 International Conference on, pp. 585-588.

[113] O. Cure, "Designing patient-oriented systems with semantic web technologies" 2003,Computer-Based Medical Systems, 2003.Proceedings.16th IEEE Symposium, pp. 195-200.

[114] Hai Zhuge, "Semantic grid: scientific issues, infrastructure, and methodology" Communications of the ACM vol. 48, no. 4, pp. 117-119, Apr.2005.

[115] Carole Goble and David De Roure, "The Semantic Grid: Myth Busting and Bridge Building" 2004,16th Eur.conf.Artificial Intelligence (ECAI-2004).

[116] David Piggott, Conor Teljeur, and Alan Kelly, "Exploring the potential for using the grid to support health impact assessment modelling" Parallel Computing vol. 30, no. 9-10, pp. 1073-1091, Sept.2004.

[117] Dipak Kalra, Peter Singleton, David Ingram, Jo Milan, J.MacKay, D.Detmer, and Alan Rector, "Security and confidentiality approach for the Clinical E-Science Framework (CLEF)" 2005.

[118] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto: X.902. Modelo de referencia: Fundamentos" 1995.

[119] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.903. Modelo de referencia: Arquitectura" 1995.

[120] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.904. Modelo de referencia: Semántica arquitectural" 1997.

[121] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.910. Marco de denominación" 1998.

[122] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.911. Modelo de referencia: Lenguaje de empresa" 2005.

[123] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.920. Lenguaje de definición de interfaz" 1997.

[124] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.930. Referencias de interfaz y vinculación" 1998.

[125] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.931. Soporte de protocolo para interacciones computacionales" 1999.

[126] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.950. Función intermediación: Especificación" 1997.

[127] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.952. Función intermediación: Provisión de la función intermediación a través del servicio de directorio de interconexión de sistemas abiertos" 1997.

[128] UIT-T.Sector de Normalización de las telecomunicaciones de la UIT, "SERIE X: REDES DE DATOS Y COMUNICACIÓN ENTRE SISTEMAS ABIERTOS. Tecnología de la información - Procesamiento distribuido abierto:X.960. Función de depositario de tipos." 1999.

[129] Jeffrey Parsons and Yair Wand, "Emancipating Instances from the Tyranny of Classes in Information Modeling" ACM Transactions on Database Systems vol. 25, no. 2, pp. 228-268, June2000.

[130] "Página Web de GALEN" 2006.
URL:www.opengalen.org/
Nota: Último acceso enero 2006

[131] "Página web de MedDRA - the Medical Dictionary for Regulatory Activities" 2006.
URL:www.meddramsso.com/NewWeb2003/index.htm
Nota: Último acceso enero 2006

[132] R. onso-Calvo, V. Maojo, H. Billhardt, F. Martin-Sanchez, M. Garcia-Remesal, and D. Perez-Rey, "An agent- and ontology-based system for integrating public gene, protein, and disease databases" Journal of Biomedical Informatics vol. In Press, Uncorrected Proof.

[133] TINA-C, "TINA Object Definition Language Manual. Version: 2.3 (Deliverable)" 1996.

[134] CEN /TC 251.Secretariat: SIS, "Health informatics - Data types" 2003.

[135] "Página web del Lenguaje Unificado de Modelado (UML)" 2006.
URL:www.uml.org/
Nota: Último acceso enero 2006

[136] K. Bernstein, M. Bruun-Rasmussen, S. Vingtoft, S. K. Andersen, and C. Nohr, "Modelling and implementing electronic health records in Denmark" International Journal of Medical Informatics vol. 74, no. 2-4, pp. 213-220, Mar.2005.

[137] Gary A.Ham, "An Integration Architecture for Implementation of the DoD Computer-based Patient Record" 2005.

[138] R. Neame and M. J. Olson, "Security issues arising in establishing a regional health information infrastructure" International Journal of Medical Informatics vol. 73, no. 3, pp. 285-290, Mar.2004.

[139] M. A. Hoffman, "The genome-enabled electronic medical record" Journal of Biomedical Informatics vol. In Press, Uncorrected Proof.

[140] M. W. A. Steen and J. Derrick, "ODP enterprise viewpoint specification" Computer Standards & Interfaces vol. 22, no. 3, pp. 165-189, Aug.2000.

[141] B. Blobel, "Authorisation and access control for electronic health record systems" International Journal of Medical Informatics vol. 73, no. 3, pp. 251-257, Mar.2004.

[142] Carlos Hernández Salvador, "Modelo de Historia Clínica Electrónica para Teleconsulta Médica." 2004.
Notes: Tesis Doctoral

[143] "Ontología publicada en la universidad de Standford y usada por miembros del openEHR" 2005.
URL:http://protege.stanford.edu/ontologies/ontologies.html
Nota: Última revisión parcial enero 2005

[144] "Página de JESS (he Rule Engine for the JavaTM Platform)" 2006.
URL:herzberg.ca.sandia.gov/jess/
Nota: Último acceso enero 2006

[145] Isabel Román, Laura M.Roa, Germán Madinabeitia, and Ramón García, "OPEN LABORATORY INFORMATION SYSTEM BASED ON CEN DATA TYPES AND LOINC CODES" 3rd European Medial & Biological Engineering Conference IFMBE European Conference on Biomedical Engineering 2005.

[146] ITU-T, "Rec. X.950-Information technology - Open distributed processing - Trading function: Specification" 1997.

[147] ITU-T, "Rec. X.960-Information Technology . Open Distributed Processing . Type Repository Function" 1999.

[148] Rama Akkiraju, Joel Farrell, John Miller, Meenakshi Nagarajan, Marc-Thomas Schmidt, Amit Sheth, and Kunal Verma, "Web Service Semantics -WSDL. Technical Note Version 1.0. Ajoint UGA-IBM technical Note" 2005.
URL:lsdis.cs.uga.edu/projects/METEOR-S/WSDL-S

[149] Jos de Bruijn, Holger Lausen, Reto Krummenacher, Axel Polleres, Livia Predoiu, Michael Kifer, and Dieter Fensel, "D16.1v0.21 The Web Service Modeling Language WSML final Draft" 2005.
URL:www.wsmo.org/TR/d16/d16.1/v0.21
Nota: WSMO. Último acceso noviembre 2005

[150] "Página del pluggin Ontoviz para Protégé (Visualizing Protégé Ontologies)" 2006.
URL:http://protege.stanford.edu/plugins/ontoviz/ontoviz.html
Nota: Último acceso enero 2006

[151] Matthew Horridge, "A Practical Guide To Building OWL Ontologies With The Protégé-OWL Plugin. Edition 1.0" 2004.

[152] "Página web de Graphviz (Graph Visualization Software)" 2006.
URL:www.research.att.com/sw/tools/graphviz/
Nota: Último acceso enero 2006

[153] E. Babulak, "Quality of service provision assessment in the healthcare information and telecommunications infrastructures" International Journal of Medical Informatics vol. 75, no. 3-4, pp. 246-252, 2006.

[154] B. Bordbar, J. Derrick, and G. Waters, "Using UML to specify QoS constraints in ODP" Computer Networks vol. 40, no. 2, pp. 279-304, Oct.2002.

[155] "Página web de JENA (A Semantic Web Framework for Java)" 2006.
URL:jena.sourceforge.net/
Nota: Último acceso enero 2006

[156] "Página web del proyecto Protègè" 2006.
URL:protege.stanford.edu/
Nota: Último acceso enero 2006

[157] "Security Assertion Markup Language (SAML)2.0 Technical Overview. Draft 01" 2005.
Nota: Último acceso septiembre 2005