

Servicios Web como soporte de investigación de las redes de investigación en biomedicina

Òscar Coltell*¹, Dolores Corella**, Ricardo Chalmeta*

*: Grupo de Integración y Re-Ingeniería de Sistemas (IRIS). Departamento de Lenguajes y Sistemas Informáticos. Universitat Jaume I, Castellón.

** : Unidad de investigación en Epidemiología Genómica y Molecular. Departamento de Medicina Preventiva y Salud Pública. Universitat de Valencia. Valencia

Resumen

Objetivo: El Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+I) para el periodo 2003-2006, en el área de Biomedicina, establece un marco de financiación basado en redes de investigación, con la participación de centros, grupos, institutos y otros organismos y empresas privadas. Dada la distribución geográfica de los participantes en las redes y la heterogeneidad de los mismos, los mecanismos de coordinación y comunicación son vitales para el buen funcionamiento de las redes. Y como soporte de la comunicación es necesario disponer de tecnologías de Internet que permitan la comunicación, el intercambio de información, compartir los resultados, proteger la información y mantener las políticas de difusión. Por lo tanto, el objetivo de este trabajo es presentar una propuesta de un Modelo de Servicios Web para el soporte de la investigación biomédica y la coordinación de redes de investigación en Biomedicina.

Metodología: En el diseño del Modelo de Servicios Web se aplican los principios, normas, metodologías, técnicas y herramientas de la Ingeniería del Software de Sistemas Web. Concretamente, se utiliza en el diseño la notación UML y el Proceso Unificado de *Rational*. Además, se tienen en cuenta las recomendaciones de diseño ergonómico de páginas y servicios web.

Resultados: Se presenta un modelo conceptual de la integración de servicios Web como soporte de la coordinación, comunicación e investigación en redes de investigación en Biomedicina. En el modelo se contemplan los servicios de comunicación (de tipo punto a punto y difusión), los servicios de soporte a la coordinación (mecanismos de emisión de directivas y recepción de informes de los nodos de la red), y los servicios de repositorio (almacenamiento, recuperación y distribución de los resultados generados). También se incluyen las funciones de gestión y mantenimiento, y de seguridad de los recursos y protección de la información.

Discusión: La propuesta que se presenta corresponde a un subproyecto en curso dentro de una de las redes de investigación que han sido aceptadas para su financiamiento. Se trata de la red G03/140 "Efectos de la dieta tipo mediterránea en la prevención primaria de la enfermedad cardiovascular", cuyo coordinador es el Dr. Ramón Estruch, del grupo MIDI-IDIBAPS del Hospital Clínic, Universitat de Barcelona.

¹ Autor de contacto: Dpto. de Lenguajes y Sistemas Informáticos. U. Jaume I. Campus de Riu Sec, s/n. 12071-Castellón. coltell@lsi.uji.es. Tel. 964-728314. Fax. 964-728435.

Palabras Clave

Servicios Web, Ingeniería de Web, Tecnologías de Soporte de la Investigación, Modelo de Servicios Web, Biomedicina, Red de Investigación en Salud.

1. Antecedentes

En la Orden SCO/709/2002 de 22 de Marzo, el Ministerio de Sanidad y Consumo convocó la concesión de ayudas para el desarrollo de Redes Temáticas de Investigación Cooperativa, dentro del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+I) para el periodo 2003-2006, en el área de Biomedicina. Esta concesión se atenía a las directrices del VI Programa Marco de la Unión Europea. Este Plan establece un marco de financiación basado en redes de investigación, con la participación de centros, grupos, institutos y otros organismos y empresas privadas.

Posteriormente, en la Resolución de 30 de diciembre de 2002 de la Dirección del Instituto de Salud "Carlos III", se aprueba la concesión y denegación de ayudas para el desarrollo de Redes Temáticas de Investigación Cooperativa. En resumen, las ayudas concedidas en su primera anualidad han sido para 56 del total de solicitudes presentadas. Los autores pertenecen a grupos de investigación que forman parte de dos de estas redes financiadas. Una de ellas es la denominada "INBIOMED. Plataforma de almacenamiento, integración y análisis de datos clínicos, genéticos, epidemiológicos e imágenes orientada a la investigación sobre patologías", con el código FIS G03/160; y la otra es la denominada "Efectos de la dieta tipo mediterránea en la prevención primaria de la enfermedad cardiovascular", con el código FIS G03/140.

Esto es un indicador de la tendencia actual en investigación que ya lleva bastante tiempo en otros países: la investigación cooperativa. Uno de los principales exponentes es el Proyecto del Genoma Humano (HUGE en inglés), que ha exigido una enorme cantidad de esfuerzos científicos y tecnológicos y cuyo éxito no hubiera sido posible sin la investigación y desarrollo cooperativos de muchos grupos y entidades de investigación, universidades, empresas y estamentos oficiales (Backofen et al., 2001).

En España, con respecto a las Redes Temáticas de Investigación Cooperativa, que determinan la distribución geográfica de los participantes en las redes y la heterogeneidad de los mismos, los mecanismos de coordinación y comunicación serán vitales para el buen funcionamiento de las redes. Y como soporte de la comunicación es necesario disponer de tecnologías de Internet (e Internet V2) que permitan la comunicación, el intercambio de información, compartir los resultados, proteger la información y mantener las políticas de difusión de resultados.

Por otra parte, el perfil profesional, técnico y científico medio de la mayoría del personal implicado en estas redes de investigación está principalmente caracterizado por médicos clínicos, médicos analistas, farmacéuticos, biólogos moleculares, bioquímicos, epidemiólogos, y otros profesionales que no tienen preparación específica en Tecnologías de Información y Comunicación (TIC). Estos investigadores dominan su medio particular y solamente necesitan las TIC como instrumento para el manejo e intercambio de la información generada en su dominio. Es decir, en esencia los investigadores de las redes son productores y consumidores de información científica, pero el almacenamiento y recuperación, intercambio, proceso y protección de esta información deben ser unos servicios proporcionados por las TIC. Si cada investigador o grupo de investigadores trabaja en una localización geográfica distinta, entonces el contacto personal y el intercambio físico es menos frecuente y los servicios mencionados deben estar asumidos por tecnologías que faciliten el trabajo cooperativo remoto. En resumen, se está hablando del concepto de Servicios Web.

Por lo tanto, el **objetivo** de este trabajo es presentar una propuesta de un *Modelo de Servicios Web* para el soporte de la investigación biomédica y la coordinación de redes de investigación en Biomedicina

En la sección 2 se presenta la metodología y los enfoques tecnológicos aplicados en el trabajo. En la sección 3 se describen las características generales de la investigación cooperativa y se expone un modelo de coordinación e intercambio de información. En la sección 4 se estudian los criterios que deben regir en el diseño de un modelo de servicios Web y se presentan las líneas generales del modelo. Y en la sección 5 se expresan las conclusiones obtenidas.

2. Metodología

En el diseño del Modelo de Servicios Web se aplican los principios, normas, metodologías, técnicas y herramientas de la Ingeniería del Software de Sistemas Web (Pressman, 2001) (Sommerville, 2002). Concretamente, se utiliza en el diseño la notación UML y el Proceso Unificado de *Rational* (Jacobson et al., 1999). Además, se tienen en cuenta las recomendaciones de diseño ergonómico de páginas y servicios web.

UML es un lenguaje abierto y adecuado para otros ámbitos. Es decir, que fue creado como un lenguaje de representación de sistemas software orientado a objetos (<http://www.uml.org>), pero permite nuevas incorporaciones de elementos de representación para distintos ámbitos, denominados *artefactos*, para poder modelar otro tipo de sistemas, como es el caso del que aquí se describe. Por otra parte, este lenguaje ya es un estándar OMG (*Object Management Group*) desde 1997 con la versión 1.3 (<http://www.omg.org>). Además, se está trabajando actualmente en el estándar UML2 también en el marco de OMG (<http://www.uml.org>). Por otra parte, UML ha sido adoptado tácitamente por la industria de sistemas informáticos que, además, ha sido uno de los grandes impulsores de su evolución y expansión. Todo ello justifica la elección de dicho lenguaje como técnica estándar de modelado.

En el proceso de concepción y diseño del modelo también se han tenido en cuenta consideraciones generales sobre la aplicación de tecnologías de Internet en el ámbito de la salud (Jones, 1998) que es donde se mueven la mayoría de las redes subvencionadas, los aspectos de seguridad y privacidad de la información biomédica (Rindfleisch, 1997), (Cass et al., 2002), y también los aspectos éticos y sociales (Sackman, 1997).

El concepto de servicio Web está muy relacionado con las tecnologías de comercio electrónico (Sens, 2002) (Papazoglou, 2003), donde se estudian con mucho rigor los aspectos de seguridad en los servicios y las transacciones (Selkirk, 2001) (Weippl et al., 2003), y los aspectos de rendimiento (Keller, 2003) y capacidad de servicio con plataformas heterogéneas (Shah, 2003) y asociación con las aplicaciones que soportan los servicios (Eggert, 1999). Además de los conocimientos teórico prácticos sobre servicios Web que se manejan en la actualidad (IBM developerWorks, 2003) y del espectro de tecnologías de soporte existentes (Tsalgatidou et al., 2002), existe en la actualidad un conjunto de *Web Services Working Groups* en el consorcio WC3 que trabajan para sistematizar y estandarizar dichos servicios en aspectos como la arquitectura de servicios (W3C[A], 2003), el lenguaje de descripción (W3C[B], 2003), y los requisitos de diseño (W3C[C], 2003). Por otra parte, se ha tenido en cuenta las tecnologías GRID para el ámbito científico que introducen el nuevo concepto de servicios Web de GRID (Gannon et al., 2002) (Silva, 2003).

3. Caracterización de la investigación cooperativa

La investigación cooperativa se caracteriza por la colaboración de numerosos grupos, que presentan distintos perfiles científicos, están diseminados geográficamente, y tienen intereses diferenciados, pero supeditan su actuación a los objetivos generales del proyecto que todos llevan a cabo. En la investigación cooperativa en Biomedicina, los grupos colaboradores deben intercambiar información y, en ocasiones puntuales, personal investigador, y además deben intercambiar muestras biológicas. Por lo tanto, es necesario establecer los distintos canales de intercambio y los mecanismos para efectuar los intercambios.

Para ello es necesario diseñar un modelo de trabajo cooperativo.

En el caso concreto de las Redes Temáticas de Investigación Cooperativa, la propia convocatoria fija las líneas generales del modelo mencionado: un modelo coordinado por un único nodo. Por lo tanto, en esta sección se describe un **modelo de investigación cooperativa coordinada por un único nodo de la red**. Este modelo será el que determine el *modelo de servicios web* que debe soportar las actividades de investigación cuya descripción se verá en la sección siguiente.

3.1. La Coordinación

En el modelo de investigación cooperativa coordinada de una red temática se debe distinguir un factor principal que es la *Coordinación de la Red*. Esta se basa en una arquitectura organizativa multicapa, que integra flujos y acciones en cada una de sus capas y está comandada por el *nodo coordinador* de la red. Los flujos que caracterizan la coordinación son los de las muestras biológicas y los de información. Las acciones que determinan la coordinación son la supervisión, las reuniones y el establecimiento de políticas para regular la difusión de resultados.

3.2. Caracterización de la arquitectura de coordinación

La *arquitectura de coordinación* es una estructura de gestión que está formada por varias capas, correspondientes a determinados flujos de muestras y de información, relacionadas entre sí por los mecanismos de coordinación (Figura 1). Se entiende que uno de los objetivos de la Red es el desarrollo y culminación de un proyecto de investigación que se compone de subproyectos a cargo de los distintos nodos de la Red. En la Figura 1 se muestra un ejemplo de arquitectura de coordinación donde hay nodos que pertenecen a un solo centro y hay nodos que pertenecen al mismo centro.

Uno de los circuitos básicos en la Red es el que determina la circulación de las muestras biológicas entre los centros y los nodos. Este circuito está modelado en la *Capa de Flujo de Muestras* de la arquitectura. Otro de los circuitos básicos es el que marca los flujos de información entre nodos y entre Coordinación de la Red y nodos. Este circuito está modelado por la *Capa de Flujo de Información*. La colaboración para actividades que no son exclusivamente de investigación, entre nodos de la Red o con entidades externas a la Red, está representada por la *Capa de Flujo de Colaboración*. Uno de los mecanismos de coordinación principales está basado en la realización de reuniones de varios tipos y finalidades. Otro de los mecanismos consiste en las acciones de difusión de políticas, directrices y resultados. A continuación se describen estos circuitos.

La Capa de Flujo de Muestras

El flujo de muestras biológicas se efectúa entre los centros y entre los nodos que se encargan de la investigación clínica y la investigación genómica y/o proteómica, analítica, etc. Pueden existir varios tipos de flujo en función del tipo de muestras y del estado de elaboración de las mismas. Por ejemplo, muestras de sangre de pacientes, ADN aislado, muestras para aislar ARN, etc.

Debe diseñarse un conjunto de protocolos que establezcan la secuencia de tratamientos que sufrirá cada tipo de muestra en los flujos de esta capa. Pero también deben diseñarse los mecanismos de control y seguimiento de las muestras, así como el esquema de autorizaciones para manejarlas y almacenarlas. Hay que tener en cuenta que las muestras biológicas procedentes de estudios en humanos también están sujetas a disposiciones legales y se debe tener autorización previa de los correspondientes comités de bioética de cada centro o institución que participa en la red.

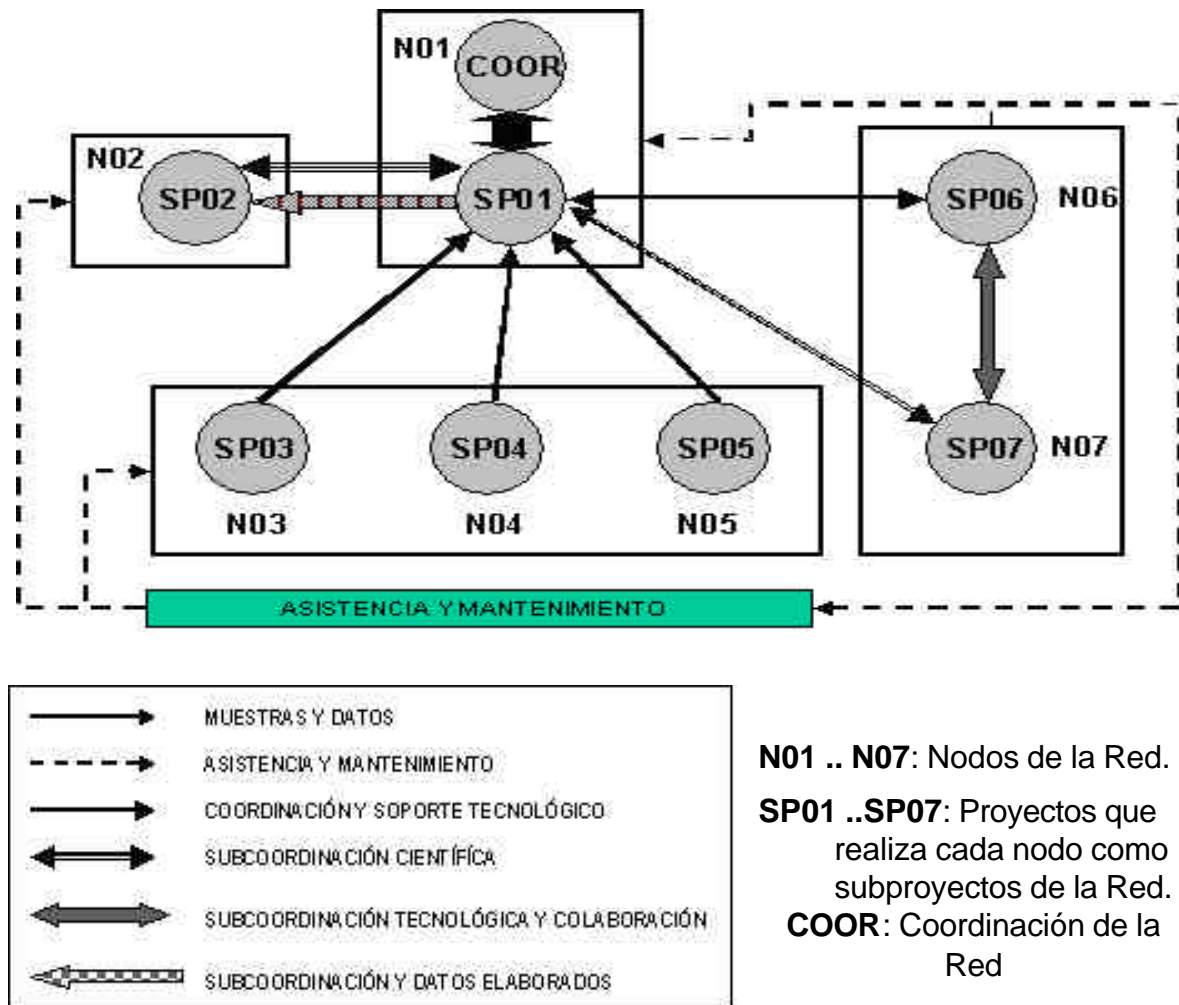


Figura 1. La arquitectura de coordinación

La Capa de Flujo de Información

La información que se manejará en la Red será fundamentalmente información de control e información de producción. La *información de control* es la que produce la Coordinación y se compone de las directrices generales para todos los nodos y subproyectos y particulares para cada nodo y/o subproyecto (Figura 1). Los medios tecnológicos que pueden soportar el flujo de la información de control son el teléfono, el fax, y tecnología de Internet.

La *información de producción* es el conjunto de informaciones científicas relacionadas con los datos de los individuos de los estudios experimentales del proyecto de la Red; las muestras biológicas (sangre, orina, ADN, etc.) que se traspasan entre los centros; los resultados de las determinaciones (Determinaciones bioquímicas básicas (DBB) parámetros hemorreológicos, coagulación y fibrinólisis y del metabolismo hidrocarbonado y lipídico, etc.); los resultados de los análisis de ADN y ARN; y los resultados de los análisis estadísticos. Además, esta información puede estar sujeta a restricciones legales y éticas si proviene de individuos humanos. Los flujos de información de producción son de dos tipos que se describen a continuación:

- La *información de producción de tipo A* es información común para las hipótesis generales del proyecto general. Corresponde al conjunto de datos generados por cada uno de los subproyectos que realizan la investigación clínica o básica específica.

- La *información de producción de tipo B* es información específica para tratar las hipótesis concretas de cada subproyecto. Es el conjunto de datos al que puede acceder solamente el subproyecto que tiene los privilegios concedidos. Normalmente, se trata de los datos producidos por el propio subproyecto y algunos conjuntos de datos adicionales. Hay que tener en cuenta que, dada la naturaleza de la información de tipo B, ésta puede estar sujeta a leyes de protección de la información y/o normas éticas
- La *información de producción de tipo C* es información para la Coordinación de la Red. La Coordinación debe tener privilegios de acceso a conjuntos de datos de tipo B, respetando toda la normativa relativa, para poder ejercer las funciones correspondientes respecto de la supervisión de resultados parciales y de las actividades científicas de cada subproyecto

La Capa de Flujo de Colaboración

La Red puede tener la responsabilidad de llevar a cabo otras actividades aparte del proyecto de investigación, como la organización de jornadas, el contacto con otras redes y otras instituciones, etc. Además, dado el carácter específico de algunas colaboraciones externas, la supervisión ordinaria de estas colaboraciones no debe estar necesariamente centralizada en la Coordinación de la Red, sino que puede estar distribuida en los nodos en función de la similitud de la especialización de cada uno de ellos y los grupos externos. Sin embargo, la Coordinación de la Red debe encargarse de las tareas de colaboración extraordinaria cuando figura la Red como única representación autorizada. Así, la *Capa de Flujo de Colaboración* se compone de cuatro tipos de flujos: flujo de coordinación interna, flujo de coordinación externa, flujo de colaboración interna, y flujo de colaboración externa. El significado de cada uno de ellos es el siguiente:

- *Flujo de coordinación interna*: Es el que se produce desde la Coordinación General hacia los tres subgrupos, en forma de directrices comunes y específicas por grupo, paso de información a cada subgrupo y establecimiento de hitos temporales para las distintas fases y actividades conforme a la planificación del proyecto. Además, el flujo se completa con el retorno de cada uno de los grupos en forma de resultados experimentales, indicadores de operación, peticiones y sugerencias.
- *Flujo de coordinación externa*: Es el que se produce desde la Coordinación General hacia los grupos externos. Se compone de tantos subflujos como número de grupos externos haya. Cada uno de estos subflujos estará compuesto a su vez del establecimiento de convenios (formales o informales) de colaboración mutua supeditada al objetivo del proyecto, del intercambio de información de estructura y funcionamiento, del intercambio de resultados experimentales y, si procede, del intercambio de muestras para aplicar determinados análisis comparativos o complementarios.
- *Flujo de colaboración interna*: Es el que se produce entre pares de subgrupos dentro del funcionamiento ordinario del grupo, poniendo en común los resultados parciales de las actividades y tareas, intercambiando información y, en el caso del Subgrupo de Bioinformática, asesorando y asistiendo técnicamente a cada uno de los otros subgrupos.
- *Flujo de colaboración externa*: Es el que se produce entre cada uno de los subgrupos y su respectivo subconjunto de grupos externos, dentro del funcionamiento ordinario del proyecto. En este flujo se realizan las cláusulas de cada uno de los convenios firmados, formal o informalmente, con cada uno de los grupos externos. Las tareas concretas a desempeñar dentro de este flujo serán las siguientes: el intercambio de información sobre protocolos experimentales, el intercambio de resultados experimentales y, si procede, del intercambio de muestras para aplicar determinados análisis comparativos o complementarios, y el asesoramiento específico en el diseño experimental y en la interpretación de los resultados obtenidos. Esto último se producirá principalmente desde los grupos externos.

El Mecanismo de Relaciones Interproyecto

El *mecanismo de relaciones interproyecto* es un mecanismo de coordinación que interactúa con las capas de Flujo de Muestras y Flujo de Información. Fundamentalmente se basa en tres tipos de Acciones de Coordinación: Reuniones Personales, Reuniones Virtuales y Conexiones Punto a Punto.

Las *Reuniones Personales* son elementos imprescindibles en cualquier ámbito en que deba ponerse en funcionamiento un mecanismo de coordinación. Habrá dos tipos: Generales y Particulares. En cada uno de estos tipos de reuniones se expone la finalidad, la periodicidad, los asistentes y los resultados.

Las *Reuniones Virtuales* se realizan para complementar las reuniones personales y para minimizar los problemas de agenda de los investigadores. Se convocan cuando surja la necesidad, sin que deban ser programadas, mediante el uso de los soportes tecnológicos como lista de correo, teléfono o fax. La ventaja de este tipo de reuniones es que cada uno de los integrantes del Proyecto Coordinado puede participar para expresar sus opiniones sin restricciones y sin problemas de desplazamientos.

Las *Conexiones Punto a Punto* engloban los tipos de comunicaciones habituales con tecnologías y medios consolidados, tales como el teléfono, el fax y los envíos de documentación y muestras por mensajería. Estas acciones sirven para resolver incidencias puntuales que exijan una resolución inmediata sin necesidad de establecer ni objetivos, ni resultados a priori.

El Mecanismo de Políticas y Responsabilidades

El *Mecanismo de Políticas y Responsabilidades* es un mecanismo de coordinación que interactúa principalmente con la Capa de Flujo de Información y parcialmente con el Mecanismo de Relaciones Interproyecto. Fundamentalmente se basa en dos tipos de Acciones: establecimiento de las políticas y responsabilidades de las actividades de investigación; y establecimiento de las políticas y responsabilidades de las actividades de difusión de resultados.

El Mecanismo de Coordinación Tecnológica

El *Mecanismo de Coordinación Tecnológica* es un mecanismo de coordinación que implica principalmente a la Coordinación y a los nodos o equipos de soporte tecnológico de la Red, si los hubiere. Estos equipos están encargados de diseñar, desarrollar e implantar los soportes tecnológicos necesarios para poder llevar a cabo la coordinación con todas las garantías de fiabilidad, seguridad y calidad. Por otra parte, deben actuar como apoyo metodológico y tecnológico a petición de la Coordinación, y prestar asistencia técnica al resto de nodos.

4. Descripción de los servicios Web

El modelo descrito en la sección anterior requiere de soporte tecnológico avanzado con la aplicación de Servicios Web y técnicas, como *Groupware*, para la gestión, supervisión y coordinación de las actividades de cada uno de los nodos de la Red. En esta sección se exponen las líneas generales de un *Modelo de Servicios Web* que actúe de soporte tecnológico al *modelo de investigación cooperativa coordinada por un único nodo de la red*.

En primer lugar, se deben establecer las necesidades y aspectos tecnológicos y de diseño que van a determinar las especificaciones del modelo. En segundo lugar, se describe la arquitectura general del modelo con los servicios principales. A continuación se desarrollan estos dos aspectos.

4.1. Criterios tecnológicos y de diseño

Dado que las Redes Temáticas de Investigación Cooperativa no se han creado principalmente para el desarrollo de sistemas tecnológicos dentro de las TIC, sino para la investigación en Biomedicina, se debe considerar que disponen de poco personal especializado en tecnología informática. Así, los sistemas informáticos de soporte que se les proporcione deben cumplir los siguientes requisitos:

- Deben ser fáciles de usar.
- Deben presentar interfaces universales independientes de las plataformas informáticas de que pueda disponer cada nodo y cada investigador.
- Deben ser flexibles según las necesidades de uso de cada investigador y nodo.
- Deben ser fácilmente mantenibles.
- Deben proporcionar mecanismos de almacenamiento, recuperación y tratamiento básico de la información generada en la Red.
- Deben proporcionar herramientas o funciones para el tratamiento específico de la información almacenada: análisis estadísticos, análisis genómicos, etc.
- Deben proporcionar mecanismos de seguridad de los recursos, de protección de la información y de control de accesos.
- Deben proporcionar mecanismos de coordinación y comunicación entre nodos.

Por otra parte, los requisitos anteriores tienen que estar complementados por los criterios de diseño siguientes:

- La facilidad de uso se debe proporcionar mediante el diseño ergonómico de las interfaces de los distintos servicios y una arquitectura de servicios integrada.
- La universalidad de las interfaces debe estar basada en la utilización de herramientas de navegación estándar para el acceso a los servicios, como los navegadores de Internet, y la utilización de lenguajes y técnicas de diseño acordes con los principios de la Ingeniería de software de Web.
- La flexibilidad de uso según las necesidades de cada usuario se consigue con la creación y gestión de perfiles de usuario. En estos perfiles, se muestra el conjunto de funciones que interesa al usuario y se oculta el resto.
- La facilidad de mantenimiento se consigue mediante una estructura de capas donde solamente se muestran al usuario medio las funciones sencillas en las cuales ha sido formado. Y las funciones complejas se ocultan salvo para los usuarios avanzados o los administradores de los servicios. Esto también tiene que ver con la gestión de perfiles de usuario.
- Los mecanismos de almacenamiento, recuperación y tratamiento básico de la información se basan en gestores de bases de datos estándar funcionando en servidores de bases de datos. Dado que la coordinación está asumida por un solo nodo, es recomendable que dicho nodo asuma las responsabilidades tecnológicas y sea la depositaria de los sistemas informáticos como los servidores mencionados.

- Las herramientas o funciones para el tratamiento específico de la información almacenada se basan en aplicaciones, de tipo estándar o desarrolladas para la Red, funcionando en servidores de aplicaciones. El nodo coordinador también debería ser depositario de estos sistemas.
- Los mecanismos de seguridad de los recursos, siendo estos recursos los sistemas informáticos, las muestras y la información, pueden ser muy variados, pero todos deben estar sujetos a una política de seguridad común de la Red establecida previamente. Los mecanismos de seguridad más habituales son los cortafuegos para los servidores de bases de datos y aplicaciones, las copias de seguridad, los planes de contingencia, sistemas de cifrado con claves públicas (PKI), etc.
- ? Los mecanismos de protección y privacidad de la información deben estar sujetos también a las disponibilidades legales vigentes. Para ello es necesario cumplir con una serie de procedimientos de inventario de datos y ficheros de datos y la comunicación a los estamentos oficiales pertinentes. Por ejemplo, si la Red maneja datos personales de los individuos que participan en los experimentos, está sujeta a la Ley de Protección de Datos Personales, y debe comunicar a la Agencia de Protección de Datos el inventario de ficheros electrónicos de este tipo que va a manejar.
- ? Los mecanismos de control de accesos deben estar basados en una estructura de Servicios Web en dos dimensiones: vertical y horizontal. La *dimensión vertical* establece varias *capas de servicios* según categorías de usuarios. Cada categoría de usuarios tiene asignada una capacidad técnica distinta y unos privilegios de acceso y utilización diferenciados, Por tanto, cada categoría sólo puede acceder a su capa de servicios correspondiente y a las que tienen menores restricciones. Por la otra parte, la dimensión horizontal establece varias *vistas de servicios o datos* según el perfil de usuario. Dos usuarios de la misma categoría pueden establecer parcelas de información que sean exclusivas de cada uno de ellos. O también puede haber determinados *grupos de usuarios* de la misma categoría que necesiten unos servicios, y el resto de usuarios no deban acceder a los mismos.
- ? Los mecanismos de coordinación y comunicación entre nodos deben estar apoyados por determinadas tecnologías y mecanismos de Internet y Web, como los servicios de Chat (IRC), las listas de correo, los tableros de anuncios, videoconferencia, etc., todo ello soportado por servidores de web que, además estén conectados a los servidores de bases de datos y de aplicaciones.

Estos requisitos son generales para el Modelo de Servicios de Red. Sin embargo, en cada Red deben asignarse pesos y prioridades en función de sus intereses particulares y su configuración individual, pero sin dejar nunca de lado los requisitos mínimos de seguridad y fiabilidad.

4.2. Categorías y perfiles de usuarios

En la subsección anterior se han mencionado el concepto de categorías de usuarios asociadas a los mecanismos de control de accesos, y el concepto de perfiles de usuario con relación a la flexibilidad de uso y al control de acceso. Las categorías de usuarios son las siguientes:

- ? Público en general: son usuarios externos a la Red. No pueden modificar nada que esté fuera del área pública.
- ? Usuarios ordinarios: son todos los investigadores adscritos a la Red.
- ? Usuarios avanzados: son los investigadores adscritos a la Red que tienen mayores privilegios que los usuarios avanzados porque tienen determinadas responsabilidades de gestión y control.
- ? Coordinación de la Red: es un usuario especial reservado únicamente a la persona que ejerza la función de coordinación de la Red.
- ? Administradores de servicios y sistemas: son los usuarios con privilegios de administración de los distintos servicios y servidores que componen el sistema global. Por ejemplo, debe existir un administrador de base de datos, un administrador de seguridad, un administrador del servidor web, un administrador de aplicaciones, etc.

Dentro de cada una de estas categorías se pueden establecer grupos de usuarios como, por ejemplo, los investigadores adscritos a un nodo, los investigadores clínicos, los investigadores básicos, etc. Cada uno de estos grupos tendrá un gestor, que será un usuario avanzado, para proceder a la inclusión y exclusión de un usuario a su grupo.

La gestión de perfiles de usuario permite definir usuarios individuales dentro de cada categoría de usuarios y asignarles todos o parte de los privilegios de acceso correspondientes.

Según la envergadura y cantidad de los servicios Web ofertados y la disponibilidad de personal técnico, algunos perfiles de administrador pueden ser asumidos por una sola persona manteniendo la siguiente incompatibilidad: el administrador de seguridad no puede asumir la administración de bases de datos o de aplicaciones.

4.3. Arquitectura de servicios

Los servicios Web deben estar inscritos dentro de un sistema integrado de servicios Web soportado por diversas tecnologías de Internet con determinados elementos físicos. Las recomendaciones de seguridad de tecnologías de Internet y Web recomiendan que los servidores conectados a una red estén representados por distintos sistemas informáticos en función de la naturaleza de cada servidor. Así, el servidor de web debe estar en un sistema informático que no asuma ninguna función más y, además, protegido del acceso exterior por al menos un cortafuegos. A este servidor de web, que ejerce como *front-end* del sistema integrado de servicios, se conectan tantos servidores de bases de datos y aplicaciones como sean necesarios. Todos los usuarios, salvo los administradores específicos, acceden a los servicios a través del servidor de web que actúa de interfaz común entre ellos (Figura 2)

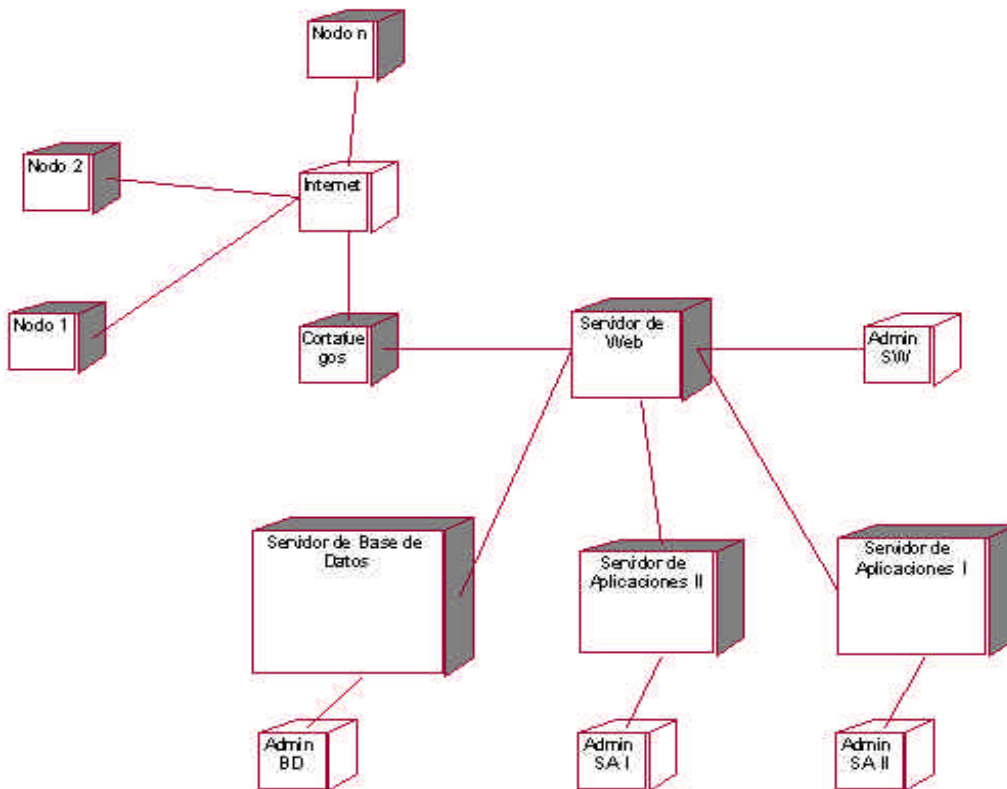


Figura 2. Arquitectura física del sistema integrado de Servicios Web (modelo de despliegue con UML)

El enfoque funcional del sistema, que ejecuta los flujos y los mecanismos descritos en el modelo de investigación cooperativa coordinada, se basa en un *marco transaccional de procesos*. Es decir, cada capa de flujo y cada uno de los mecanismos se expresa dinámicamente mediante un conjunto de procesos que da lugar en el sistema a una transacción entre usuario y uno o más servicios. De este modo, el flujo de trabajo iniciado en una transacción debe estar dirigido y apoyado por dos tipos de protocolos: el protocolo de coordinación y el protocolo de proceso.

El *protocolo de coordinación* consiste en la coordinación de las operaciones de los servicios Web entre las distintas aplicaciones, distribuidas o no, que ejecutan dichos servicios y que pueden ser bastante heterogéneas. El *protocolo de proceso* consiste en las directrices para capturar la información e intercambio de requisitos entre los distintos actores de una transacción, identificando parámetros temporales, de secuencia de acciones, y objetivos de cada una de las colaboraciones que se producen en la transacción

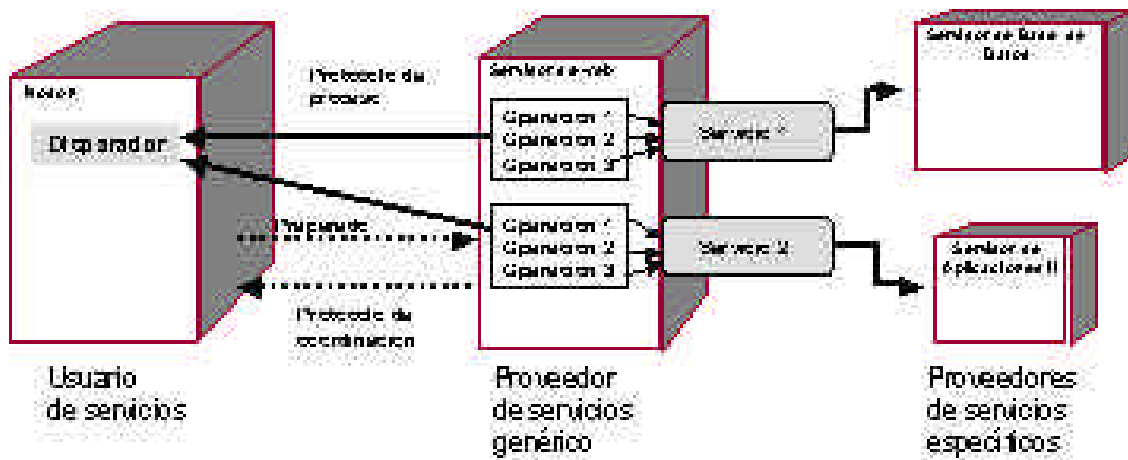


Figura 3. Esquema de una transacción en el sistema integrado de servicios

En este modelo deben incorporarse todos los mecanismos básicos asociados a los sistemas transaccionales, como los sistemas de traza, los sistemas de recuperación de transacciones, los mecanismos de *rollback*, el no repudio, etc.

Por otra parte, en el modelo se define el concepto de *transacción atómica* como el conjunto mínimo de actividades de los servicios que deben tener una salida única, sea la obtención de un resultado o la cancelación, y que no puede ser dividida en subconjuntos de actividades bajo la misma condición. De esta forma, las transacciones ordinarias que soportan los procesos de la Red pueden ser transacciones atómicas o composiciones de transacciones atómicas.

La descripción detallada de este modelo exigiría una extensión mucho mayor del documento que no es pertinente en este trabajo, donde solamente se exponen las líneas generales. La propuesta que se presenta corresponde a un subproyecto en curso dentro de la red G03/140 "Efectos de la dieta tipo mediterránea en la prevención primaria de la enfermedad cardiovascular", cuyo coordinador es el Dr. Ramón Estruch, del grupo MIDI-IDIBAPS del Hospital Clínic, Universitat de Barcelona. Uno de los primeros resultado del proyecto, todavía no concluso, será la descripción completa y pormenorizada del modelo mencionado en esta sección.

5. Conclusiones

Se presenta en este trabajo una propuesta de un Modelo de Servicios Web para el soporte de la investigación biomédica y la coordinación de redes de investigación en Biomedicina. El modelo representa un sistema integrado de servicios Web que dará soporte a las actividades de las Redes Temáticas de Investigación Cooperativa, ejecutadas según un modelo de investigación cooperativa coordinada por un único nodo que también se describe en el trabajo.

La propuesta es uno de los primeros resultados parciales de un subproyecto desarrollado en el seno de una de las redes subvencionadas dentro del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+I) para el periodo 2003-2006, en el área de Biomedicina. Esta red se denomina, "Efectos de la dieta tipo mediterránea en la prevención primaria de la enfermedad cardiovascular". Pero además, se cuenta con el apoyo metodológico y tecnológico que puede proporcionar otra de las redes subvencionadas, denominada "INBIOMED Plataforma de almacenamiento, integración y análisis de datos clínicos, genéticos, epidemiológicos e imágenes orientada a la investigación sobre patologías", siguiendo el espíritu de servicio a otras redes que animó la creación de la misma.

Agradecimientos

Este trabajo ha contado con la financiación parcial de la Red Temática de Investigación denominada "INBIOMED Plataforma de almacenamiento, integración y análisis de datos clínicos, genéticos, epidemiológicos e imágenes orientada a la investigación sobre patologías". (Código del Fondo de Investigaciones Sanitarias, Ministerio de Sanidad, G03/160), cuyo coordinador es el Dr. Fernando Martín, del Instituto de Salud Carlos III. También se ha contado con la financiación de la Red Temática de Investigación denominada "Efectos de la dieta tipo mediterránea en la prevención primaria de la enfermedad cardiovascular" (Código del Fondo de Investigaciones Sanitarias, Ministerio de Sanidad, G03/140), cuyo coordinador es el Dr. Ramón Estruch, del grupo MIDI-IDIBAPS del Hospital Clínic, Universitat de Barcelona.

Bibliografía

- Backofen R., Gilbert D. "Bioinformatics and Constraints". *Constraints*, 6; 2001: 141-156.
- Booch G, Rumbaugh J, Jacobson I. El Lenguaje Unificado de Modelado. Addison-Wesley, Madrid, 1999.
- Cass S., Riezenmann M.J. "Improving Security, Preserving Privacy". *IEEE Spectrum*, Jan.; 2002: 44-49.
- Eggert L. "Application-level differentiated services for Web servers". *World Wide Web* 2(3); 1999: 133-142.
- Gannon D., Bramley R., Fox G., Smallen S., Rossi A. , et al. "Programming the Grid: Distributed Software Components, P2P and Grid Web Services for Scientific Applications". *Cluster Computing* 5(3); 2002: 325-336.
- IBM developerWorks. "Web services specifications". IBM developerWorks. January 2003. <http://www-106.ibm.com/developerworks/webservices/library/ws-spec.html>
- Instituto de Salud Carlos III. Resolución de 30 de diciembre de 2002 de la Dirección del Instituto de Salud «Carlos III» por la que se aprueba la concesión y denegación de ayudas para el desarrollo de redes temáticas de investigación cooperativa. 3/01/2003.

Jacobson I. Booch G, Rumbaugh J. The Unified Software Development Process. Addison-Wesley, Reading, MA (USA), 1999.

Jones R.L. «The Internet and Healthcare Information Systems: How Safe Will Patient Data Be?». *IS Audit & Control Journal*, I; 1998: 25-30.

Keller A., Ludwig H. “The WSLA Framework: Specifying and Monitoring Service Level Agreements for Web Services”. *Journal of Network and Systems Management* 11(1); 2003: 57-81.

Orden SCO/709/2002 de 22 de Marzo, el Ministerio de Sanidad y Consumo convocó la concesión de ayudas para el desarrollo de Redes Temáticas de Investigación Cooperativa. BOE núm. 80, de 2 de abril de 2002.

Papazoglou M. P. “Web Services and Business Transactions”. *World Wide Web* 6(1); 2003: 49-91.

Pressman R. S. *Ingeniería del Software. Un enfoque práctico* (5ª ed.), Mc Graw-Hill, Nueva York, 2001.

Rindfleisch T.C. «Privacy, Information Technology, and Health Care». *Communications of the ACM*, 40-8; 1997: 93-100.

Sackman H. *Biomedical Information Technology. Global Social Responsibilities for the Democratic Age*. Academic Press, San Diego, CA (USA), 1997.

Selkirk A. “Using XML Security Mechanisms”. *BT Technology Journal* 19(3); 2001: 35-43.

Sens T. “Next Generation of Unified Communication for Enterprise”. *Alcatel Telecommunications Review - 4th Quarter 2002*: 1-8.

Shah U. J. “Distribute a rich user experience using Web services and a balanced computing model. Shared processing on many devices eases server load, supports multiple platforms, and delivers value to end users”. March 7, 2003. <http://www-106.ibm.com/developerworks/webservices/library/ws-balcom.html>

Silva V. “Grid Job submission using the Java CoG Kit. Writing Java-enabled Grid applications”. IBM developerWorks. February 2003. <http://www-106.ibm.com/developerworks/webservices/library/ws-gridcog.html>

Sommerville I. *Ingeniería de software*, 6ª edición, Prentice Hall – Pearson Educación, México, 2002.

Tsalgatidou A., Pilioura T. “An Overview of Standards and Related Technology in Web Services”. *Distributed and Parallel Databases* 12(2); 2002: 135-162.

W3C[A]. The Web Services Description Working Group. “First Working Draft of the Web Services Architecture”. Working Draft 2002-11-14. <http://www.w3.org/2002/ws/>

W3C[B]. The Web Services Description Working Group. “The Web Services Description Language (WSDL) Version 1.2” Working Draft 2003-03-03. <http://www.w3.org/2002/ws/>.

W3C[C]. The Web Services Description Working Group. “Web Services Description Requirements” Working Draft 2002-10-28. <http://www.w3.org/2002/ws/>

Weippl E., Essmayr W. “Personal Trusted Devices for Web Services: Revisiting Multilevel Security”. *Mobile Networks and Applications* 8(2); 2003: 151-157.